Agonists of the transcriptional Keap1-Nrf2 gatekeeper

Gacesa, Ranko

Awarding institution:
King's College London

Agonists of the transcriptional Keap1-Nrf2 gatekeeper

THESIS *incorporating publications*

Submitted in candidature for the degree of

DOCTOR OF PHILOSOPHY

from

KING'S COLLEGE LONDON

by

Ranko Gacesa

King's College London,
Faculty of Life Sciences and Medicine,
150 Stamford Street,
London SE1 9NH

Acknowledgements

Very special thanks to my supervisor Prof Paul F Long, King's College London, for advice and direction during this project. Thank you to Dr Walter Dunlap, Dr David Barlow and Professor Daslav Hranueli for their support and advice, and to Liam Doonan for his assistance in proofreading the thesis.

Thank you to Dr Antonio Starcevic, Dr Jurica Zucko and all members of research team at Section of Bioinformatics, Faculty of Food Technology and Biotechnology, University of Zagreb.

Thank you to Prof Dame Janet M Thornton providing me with the opportunity for sabbatical placement and to Dr Roman Laskowski and all members of the Thornton research group at the EMBL-EBI for supervision and assistance during the visit to the EMBL-EBI.

Finally, thanks to friends and family for their support.

Declaration

This thesis submitted for the degree of Ph.D. entitled "*Agonists of the transcriptional Keap1-Nrf2 gatekeeper*" is based on the work conducted by the author at the Institute of Pharmaceutical Sciences, King's College London between July 2013 and July 2017. All of the work described herein is original unless otherwise acknowledged in the text or by references. None of the presented work has been submitted for another degree in this or any other university.

Study presented in *Chapter 3: Rising levels of atmospheric oxygen and evolution of Nrf2* was performed, in part, during the research visit to Thornton group at European Bioinformatics Institute (EBI) and was co-supervised by Dr Roman A. Laskowski, EBI.

Results presented in *Chapter 4: Porphyra-334 and shinorine are antioxidants and antagonists of Keap1-Nrf2 binding* are a product of collaborations with Dr Geoffrey Wells and Nikolaos Georgakopoulos, UCL School of Pharmacy, and Prof Ernani Pinto, Faculty of Pharmaceutical Sciences, University of São Paulo, Brazil.

Results presented in *Chapter 5: Proteomics approach to study the UV-induced oxidative stress in yeast isolated from environments with high incidence of solar radiation* are a product of collaboration with Dr Malcolm Ward and Mr Ray Chung of KCL Proteomics Facility.

The work leading to this thesis and the publications contained herein was supported by the United Kingdom Medical Research Council (MRC grant G82144A to R.G. and P.F.L.) in partnership with Prof Daslav Hranueli, Dr Jurica Zucko and SemGen Ltd. (Zagreb, Croatia).

Ranko Gacesa,

April, 2016

Abstract

Oxidative stress has been associated with numerous degenerative diseases and disorders, as well as cancer and the process of ageing. In higher animals, the response to oxidative stress is largely regulated by the master transcription factor Nrf2, which controls the transcription of cytoprotective genes, and its inhibitor Keap1 which functions as a "sensor" of oxidative stress. Keap1-Nrf2 pathway is known to be conserved across vertebrates and certain invertebrates, but its evolution is yet to be described and it is currently unknown if microbes such as yeasts and bacteria possess this pathway. This thesis examines microbial genomes for evidence of Keap1-Nrf2 pathway, investigates the evolution of this pathway over geological time, and assesses the potential for activation of Nrf2-controlled cytoprotection by microbially produced small compounds.

The novel software for identification of distant homologs was developed and utilized to study the homologs of Keap1 and Nrf2 proteins in genomes of animals and microorganisms. The evolution of Keap1-Nrf2 pathway was reconstructed by phylogenetic studies, and the time-frame of evolution was calibrated using the fossil record. The existence of Keap1-Nrf2 pathway in fungi was also examined empirically by utilizing high-throughput proteomics to quantify the stress response mechanisms of an UV-tolerant yeast model. Structure based virtual screening was employed to identify microbial natural products with potential to activate human Nrf2 pathway by inhibiting the Keap1-Nrf2 binding, and the prospective *in-silico* activators of Nrf2 were tested *in vitro* by fluorescence polarisation and thermal shift assays to detect competitive inhibition of human Keap1-Nrf2 binding.

In-silico analyses identified that the Keap1-Nrf2 pathway exists in all major eukaryotic phyla, ranging from fungi to mammals, and that Nrf2 evolved under a selective pressure incurred by the rise of oxygen levels over geological time. The *in-silico* virtual screen identified the potential for competitive inhibition of Keap1-Nrf2 binding by mycosporine-like amino acids (MAAs), small compound UV-protective and antioxidant metabolites of marine microorganisms. This activity of MAAs was tested empirically, and the MAAs shinorine and porphyra-334 were confirmed to competitively inhibit the human Keap1-Nrf2 interaction *in vitro*. The results presented herein indicate that natural products of microorganisms, such as MAAs, are the prospective compound leads for the design of novel therapeutics to target activation of the human Keap1-Nrf2 pathway for treating degenerative diseases of oxidative stress, whilst avoiding the off-target effects of currently utilized Nrf2 activators.

Table of Contents

List of Figures

List of Tables

Abbreviations

4-DG: 4-deoxygadusol
β-NF: beta-naphthoflavone, 3-phenyl-1H-benzo[*f*]chromen-1-one
β-TrCP: β-transducin repeat-containing protein
AAPH: 2,2'-azobis(2-amidinopropane) dihydrochloride
ABTS: 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid), stable free radical
ACN: acetonitrile
AD: Alzheimer's disease
ADT: AutoDock Tools
AIC: Akaike information criterion
AICc: AIC corrected for finite sample sizes
AP-1: yeast activator protein 1 (Yap1) recognition element on the DNA
ARE: antioxidant response element, also known as electrophile response element (EpRE)
***BAX*:** gene encoding B-cell lymphoma 2 (BCL-2) associated protein X
BEAST: Bayesian Evolutionary Analysis Sampling Trees, phylogeny software
BHA: mixture of 2-tert-Butyl-4-methoxyphenol and 3-tert-butyl-4-methoxyphenol
BIC: Bayesian information criterion
BLAST: Basic Local Alignment Search Tool
BTB: Broad complex, Tramtrack and Bric-a-Brac domain
bZIP / bZip: basic leucine zipper
caffeic acid: 3-(3,4-Dihydroxyphenyl)-2-propenoic acid
cAMP: cyclic Adenosine monophosphate (AMP)
Cat / CAT: Catalase
CBP: cAMP response element-binding protein (CREB)-binding protein, transcription factor
CCP: cytochrome-c peroxidase
Cdc: cell division control protein
CDDO-TF(E)A: 2-Cyano-3,12-dioxooleana-1,9(11)-dien-28-oyl Trifluo(ethyl) Amide
CDDO-Im: 1[2-Cyano-3,12-dioxooleana-1,9(11)-dien-28-oyl] imidazole
CFU: colony forming unit
chlorogenic acid: 3-(3,4-dihydroxycinnamoyl)-quinic acid
ChIP-seq: chromatin immunoprecipitation DNA sequencing
CNC: cap and collar protein architecture
CoQ: Ubiquinol, Coenzyme-Q
CPD: cyclobutane pyrimidine dimer
CREB: cAMP response element binding protein
CRL^{Keap1}: Cullin3-Rbx1 E3 ubiquitin ligase - Keap1 complex
Cul3: Cullin-3
curcumin: (1E,6E)-1,7-bis(4-hydroxy-3-methoxyphenyl)-1,6-heptadiene-3,5-dione
DBD: DNA binding domain
DHQS: 3-dehydroquinate (DHQ) synthase
DHSP: Distant Homology Search Pipeline
DMF: dimethyl fumarate
DMSO: dimethyl sulfoxide
DPBS: Dulbecco's phosphate buffered saline
DPPH: 2,2-diphenyl-1-picrylhydrazyl, stable free radical
ECH: erythroid cell-derived protein with cap and collar protein (CNC) homology
EGCG: epigallocatechin gallate
ER: endoplasmic reticulum
ESI: electrospray ionization

EVS: 2-epi-5-epi-valiolone synthase
FKH: forkhead protein transcription factor
FoxO: forkhead box O transcription factor
FP: fluorescence polarization
FRAP: ferric-reducing antioxidant power
Ga: giga-annum, billion years before the present
Gclc: glutamate - cysteine ligase catalytic subunit
Gcn4: general control nondepressible protein 4
GO: gene ontology
GOE: Great oxygenation event
Gpx, GPx: glutathione peroxidase
Grx: glutaredoxin
GSH: glutathione
Gsr, GTRx: glutathione reductase
HCM: high-copy suppressor of Calmodulin, forkhead protein
HO-1: heme oxygenase 1 protein, encoded by *HO-1* gene (also known as *HMOX-1*)
HMM: hidden Markov model
HMMER: HMM-based homology searching utility
HPLC: high performance liquid chromatography
HSP: heat shock protein
IDH: isocitrate dehydrogenase
IVR: intervening region of Keap1 protein
JNK: Jun proto-oncogene, AP-1 transcription factor subunit (c-Jun) N-terminal kinase
Keap1: kelch-like ECH-Associated protein 1
KEGG: Kyoto Encyclopaedia of Genes and Genomes
KR: Kelch repeat
LCFA: long chain fatty acid
LC-MS/MS: HPLC coupled tandem mass spectrometry
Ma: mega-annum, million years before the present
MAA: Mycosporine-like amino acid
Maf: proto-oncogene c-Maf, V-maf musculoaponeurotic fibrosarcoma oncogene homolog
MALDI: matrix-assisted laser desorption/ionization
MAPK: mitogen activated protein kinase
MCMC: Markov chain Monte Carlo method
MEGA: molecular evolutionary genetics analysis software
M-Gly: mycosporine glycine
(M)FRTA: (mitochondrial) free radical theory of ageing
MS/MS: tandem mass spectrometry
MS: mass spectrometry
MSA: multiple sequence alignment
MudPIT: multidimensional protein identification technology
NAD(P)H: nicotinamide adenine dinucleotide (phosphate)
NCBI: National Centre for Biology Information
Neh: Nrf2-ECH homology domain
NES: nuclear exclusion signal
NF-κB: nuclear factor kappa-light-chain-enhancer of activated B cells
NLS: nuclear localization signal
NQ(O)R: NAD(P)H Quinone (oxido)reductase
NR: NCBI database of non-redundant protein sequences
Nrf2: nuclear factor erythroid 2-related factor 2
NRPS: nonribosomal peptide-synthetase

OD600: spectral absorbance at 600 nm
OMT: O-methyltransferase
ORAC: oxygen radical absorbance capacity assay
OxyR: protein transcriptional autoregulator of bacterial hydrogen-inducible genes
PBS: phosphate-buffered saline buffer
PD: Parkinson's disease
PDB: Protein Data Bank
PdxS/SNZ: pyridoxal 5'-phosphate synthase subunit associated with snazarus (SNZ) protein
PKA: protein kinase A
PPI: protein-protein interaction
Prx: peroxiredoxine
psi-BLAST: position specific iterative BLAST
Quercetin: 2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-4H-chromen-4-one
qPCR: quantitative polymerase chain reaction (also called real-time PCR)
Ras: protein belonging to Ras superfamily of small GTPases
Rap: Ras associated protein
Rbx1: cullin-RING-based E3 ubiquitin-protein ligase, encoded by *RBX1* gene
RNS: reactive nitrogen species
ROS: reactive oxygen (or oxidative) species
RS: reactive chemical species, such as free radicals ROS and RNS
RXRα: retinoid X receptor alpha
SFN: sulforaphane, 1-Isothiocyanato-4-methylsulfinylbutane
SILAC: stable isotope labelling with amino acids in cell culture
siRNA: small interfering RNA, sometimes known as short interfering RNA or silencing RNA
SIRT6: sirtuin-6
SKN-1: *Caenorhabditis elegans* homolog of vertebrate Nrf2 protein
SOD: superoxide dismutase
SOR: superoxide reductase
SoxR: protein transcriptional autoregulator of bacterial superoxide-inducible genes
SRSA: superoxide radical scavenging activity
tBHQ: tert-butylhydroquinone, 2-(1,1-Dimethylethyl)-1,4-benzenediol
TBS: tris–buffered saline buffer
TCP-1: T-complex protein 1 ATPase
TEAB: triethylammonium bicarbonate
TFA: triflouroacetic acid
TCEP: tris(2-carboxyethyl)phosphine hydrochloride
TLM: Taxonomy Landscape Mapper
TMT: tandem mass tag
trans-resveratrol: 3,5,4′-trihydroxy-trans-stilbene
***TP53*:** gene encoding the tumor protein p53
Trolox: water soluble analogue of vitamin E, 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid
Trx, TRX: thioredoxin
UCFS: University of California, San Francisco
UFF: universal force field
UPGMA: unweighted pair group method with arithmetic mean
UniProt: Universal Protein Resource
UV / UVR: ultraviolet radiation
YAP1: yeast Activator protein 1
YPD: yeast/peptone/dextrose media
VS: virtual screening

Chapter 1: Introduction

1.1 Foreword to Chapter 1

This chapter introduces the basic terminology of redox biology and the computational and empirical approaches used in this work. The concept of oxidative stress and basic terminology of redox biology are introduced in section 1.2, and the principal stress response pathway in animals is described in the section 1.3. The section 1.4 introduces a class of natural compounds which were identified as potential activators of animal stress response using the *in-silico* methods described in section 1.5. The section 1.6 describes the empirical approach used to evaluate the stress response mechanisms of *in-vivo* model to validate the *in-silico* studies of animal stress response pathway, while the sections 1.7 and 1.8 give a brief summary of the introduction and present the research hypotheses and goals of this project.

1.2 Oxidative stress

The concept of oxidative stress was introduced in 1985 by Helmut Seis as "a disturbance of the prooxidant – antioxidant balance in favour of the former." (1). The Seis' work and "oxidative stress" hypothesis was largely inspired by the discovery of free radicals in biological tissues (2), the discovery of superoxide dismutase proteins (3) and the inception of Harman's *Free radical theory of ageing* (4), which postulated that the process of ageing is caused by ROS-induced damage. Further studies, at a turn of the twentieth century, led to the realization that oxidative cellular damage is associated with innumerable diseases and disorders, including, but not limited to, carcinogenesis, inflammation, neurodegeneration and ageing (1,5,6). Since the eighties, the field of redox biology and medicine has seen an enormous expansion; for illustration, a PubMed database (7) search for "oxidative stress" lists over 150,000 articles published until the end of 2016, with over 70,000 published in the period 2011 – 2016.

It should be noted, however, that Harman and Seis's work was not without its critics. For example Hallywell and Gutteridge argued that an association between ROS and disease does not imply a causal relationship (8,9). It was also recognized at that time that redox processes and ROS are not necessary deleterious for biological systems, and are vital for immune response as well as parts of cellular metabolism such as production of derivatives of fatty acids and deoxyribonucleotides (5,10). Indeed, despite the decades of research and tens of thousands

of published studies in redox biology, the role of ROS in health, disease and ageing is still an active field of research with a number of controversies; these are briefly reviewed in Section 1.2.8.

1.2.1 Oxidative stress, terminology

The extreme levels of oxidants, such as hydrogen peroxide (H_2O_2) or pure oxygen (O_2), are toxic (6) and the disruption of critical cellular redox signalling mechanisms, e.g. mutation of superoxide dismutase (SOD) encoding gene, cause severe pathologies in model organisms (11). Defining oxidative stress, however, is not as simple as stating that ROS cause cellular damage leading to damage and number of pathologies unless kept in check by cellular antioxidant mechanisms. Plenty of evidence exists for the role of ROS, especially H_2O_2, in cellular signalling, functioning of the immune system and the control of cell cycle (12–14). In addition, it has been demonstrated that removal of ROS from healthy biological systems can be deleterious (15,16). Finally, terms such as *ROS* and *antioxidant* lack clear definition, and the recent recommendation by the journal *Free Radicals in Biology and Medicine* is to avoid use of these terms in favour of those more specific (17). Acknowledging these issues, here follows a brief introduction to the basic redox biology terms used in this thesis.

Free radical: Free radical is a chemical species with unpaired electron; free radicals common in biological systems include superoxide anion radicals (O_2^{-}•), highly active hydroxyl radicals (OH•), and diverse biomolecule radicals and peroxyls, for example lipid radicals and peroxyls formed during the lipid peroxidation process (R•, ROO•); this term also includes radicals not derived from oxygen such as nitric oxide radical (NO•) (5,18).

Reactive oxygen species (ROS): ROS are (reactive) products of molecular oxygen (O_2); these include free radicals such as hydroxyl (OH•), and non-radical species such as hydrogen peroxide (H_2O_2), hydroxide anion (OH^-) and peroxides of proteins, lipids, nucleic acids and other macromolecules (ROOH). Notably, terms ROS and free radicals are sometimes used interchangeably, despite the fact that some of the oxygen-derived species, e.g. H_2O_2 and OH^-, are not radicals. In addition, this term is sometimes used also to refer to nitrogen derived reactive species (RNS) such as NO•, NO_2 and ONOO (see (19–22) for examples). To add to the confusion, ROS is sometimes used as an abbreviation for *reactive oxidizing species* or *reactive oxidant species* (which would imply non-oxygen chemical species such as NO_2 and is arguably a more correct, albeit less often used, term for chemical sources of oxidative stress).

In order to avoid the confusion, terms *free radicals* and *ROS* are not used in this work in favour of the more generic term *reactive chemical species (RS).*

Reactive (chemical) species (RS): Unless the chemical entity involved is known, this work uses the term *reactive species (RS)* to refer to free radicals and other reactive, oxidizing, chemical species (such as ROS and RNS) involved in redox signalling networks and in the oxidative damage to biological systems, i.e. oxidative stress.

Oxidative stress: Originally defined by Sies as "*a disturbance of the prooxidant – antioxidant balance in favour of the former*" (1), the definition of oxidative stress has since been recast as "*an imbalance between oxidants and antioxidants, potentially leading to molecular damage*" (23) to acknowledge the difference between oxidative damage and physiological levels of RS involved in cellular redox-signalling networks (12). Recent years have seen major criticism of this term as unspecific and unclear (17). For example, oxidative stress is usually used in the context of oxidative damage leading to disease, but the imbalance between the oxidants and antioxidants is a natural occurrence in the cell cycle (24). Furthermore, a low level of oxidative stress can be beneficial for the cell, e.g. by inducing the cytoprotection via Nrf2 pathway (25). These realizations prompted the introduction of the term oxidative eustress to describe "positive" stress leading to the activation of cellular defences, as seen for example during aerobic exercise (24); in addition, it was suggested that oxidative stress should instead be defined as "a disruption of redox signalling and control", to account for redox signalling and to acknowledge the fact that oxidation and reduction reactions are not necessary in the balance in biological systems (25).

To make this distinction clear, the term "oxidative stress" is herein used to refer specifically to the disruptive stress, i.e., disruption of physiological redox signalling and control by aberrantly high levels of RS, potentially leading to molecular damage.

Antioxidant: This work uses the definition of Halliwell and Gutteridge and refers to an antioxidant as "a substance that, when present at low concentrations compared with those of an oxidiseable substrate, significantly delays or prevents oxidation of that substrate" (26). As follows from this definition, an antioxidant activity of a compound depends on an assay or the model used.

1.2.2 Nature and causes of oxidative stress

Reactive chemical species (RS) are unavoidable by-products of the oxidizing atmosphere of contemporary Earth, environmental toxins and pollutants, and the aerobic life-style of eukaryotic organisms (see Figure 1.1 for examples of RS sources). Within the cell, the mitochondria are a major producer of RS, primarily due to release of superoxide from complexes I and III of the electron transfer chain (27,28). Furthermore, mitochondrial enzymes involved in the citric acid cycle have also been associated with RS production (27,29). Aside from the mitochondria, a number of cytosolic protein families such as NADPH oxidases (NOX enzymes), xanthine oxidase, nitric oxide synthase, cyclooxygenases, cytochrome P450 enzymes, and lipoxygenases potentially produce RS, although these effects likely vary between cell and tissue types (29,30).

Numerous external sources have been associated with the production of RS in living systems. These include ionizing radiation, UV light and a wide range of chemicals often encountered as pollutants. For example, paraquat reacts to form peroxides or ozone; nitroaromatics, quinones and herbicides related to paraquat promote formation of superoxide; phenols, aminophenols and other chemicals are metabolized to radicals; iron and copper releasing compounds increase concentration of free metals in the cell, potentially causing formation of highly reactive hydroxyl radicals by Fenton reactions of H_2O_2 (19,31,32). Induction of oxidative stress is also observed as a result of exposure to heavy metals and various xenobiotics such as chlorinated compounds, radiation, metal ions and different pharmaceutical drugs, e.g. barbiturates and antitumor drugs cisplatin and doxorubicin (32,33).

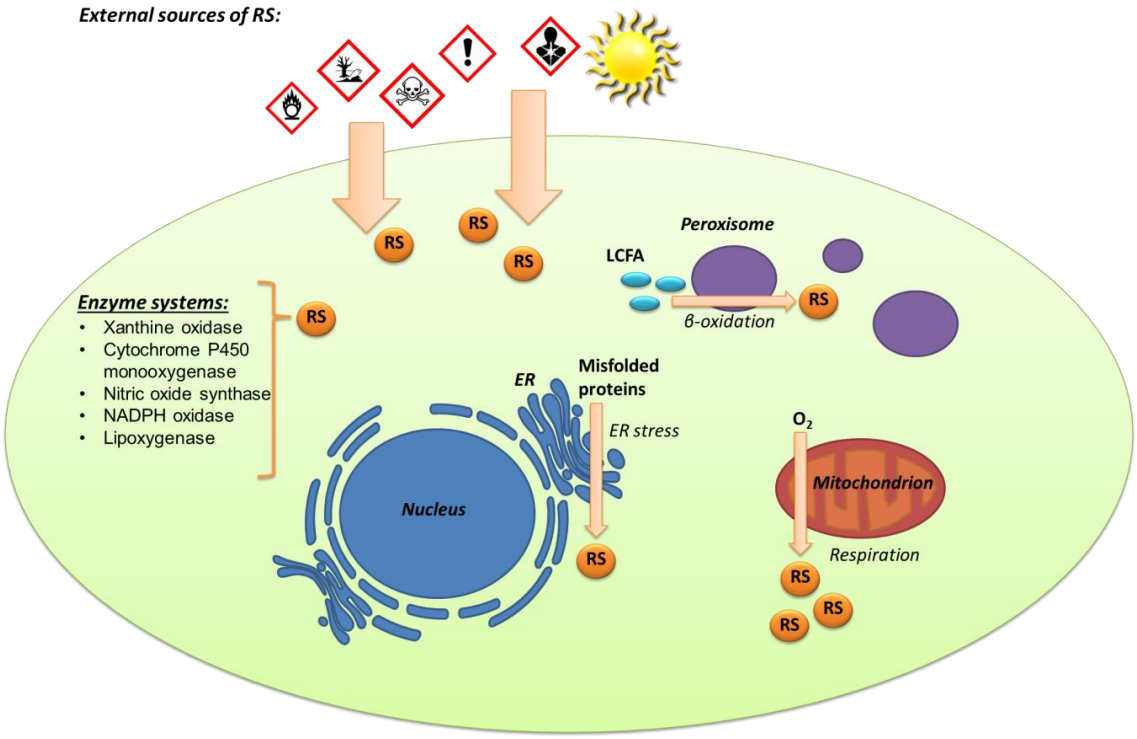


Figure 1.1 Sources of oxidative stress

External sources of RS include UV radiation, toxins, xenobiotics and various chemical stressors encountered in the environment. Internal sources of RS include mitochondrial respiratory chain, peroxisomal oxidation of long chain fatty acids (LCFA), endoplasmic reticulum (ER) stress caused by misfolded proteins and numerous enzymatic systems endogenous to the cell. Nucleus and ER are illustrated in blue, peroxisome in purple and mitochondrion in red; processes leading to formation of RS are depicted with orange arrows and RS are denoted with orange circles.

1.2.2.1 UV radiation and oxidative stress

Solar UV radiation (UVR) is a major exogenous hazard to the terrestrial and marine organisms on Earth. UVR inflicts direct damage to cellular macromolecules, e.g. DNA, and indirect, oxidative, damage by generating reactive chemical species (RS) in the cell (34). UVR is customarily divided into UV-A, UV-B and UV-C radiation based on wavelength, absorption by the atmosphere, and biological effects (35). While Earth's atmosphere and the ozone layer block short length UV-C radiation (with wavelength λ = 100 – 280 nm), they provide less protection against longer wavelength UV-B (λ = 280-315 nm) and virtually no defence against UV-A (λ = 315 – 400 nm) radiation. Ground level UVR thus comprises ~95% UV-A and ~5% UV-B, with negligible UV-C levels (36).

Therefore, while highly energetic, UV-C is of little biological significance in the natural environment as it does not reach ground level. While UV-B is a minor part of ground-level UVR (~5%), it is absorbed by DNA and causes formation of pyrimidine dimers and DNA

breaks (37). It is also absorbed by amino acid residues (mainly Trp, Tyr and Phe) and inflicts protein damage via photo-oxidation reactions (38). In addition, UV-B was also associated with induction RS and the oxidative stress (39). UV-B is the part of UV spectrum responsible for sunburn as well as sun-induced melanin production, and is the cause of direct tissue damage, mutagenic and carcinogenic effects (40). UV-A is less energetic then UV-B (due to higher wavelength) but penetrates deeper into skin tissue and has been associated with indirect damage via generation of RS and induction of oxidative stress (40).

Observational human studies and experimental studies with animal models led to the inference of a causal link between UV exposure and numerous disorders including sunburn (40), immuno-suppression (41), skin cancer (42,43) and eye diseases (44). In addition, UVR was shown also to cause degenerative changes in skin collagen, leading to accelerated ageing (45). While UV-B induced DNA damage is strongly associated with multiple variants of skin cancer, such as basal cell and squamous cell cancers, it is shown that malignant melanoma lack mutations specific to UV-B irradiation, and are likely caused by indirect effects of UV-induced oxidative stress (46,47).

1.2.3 Deleterious effects of oxidative stress

While in low concentrations, reactive chemical species (RS) are a part of cellular metabolism and signalling (13), and are kept in check by antioxidant and detoxifying mechanisms (48). Imbalance between formation and cleansing of RS however leads to elevated RS concentration and potential for cell damage – a state called oxidative stress (1). When oxidative stress occurs, RS cause cellular damage by interacting with nucleic acids, lipids and proteins (49), but also induce redox signalling that leads to upregulation of cellular protective systems to purge

excessive RS (

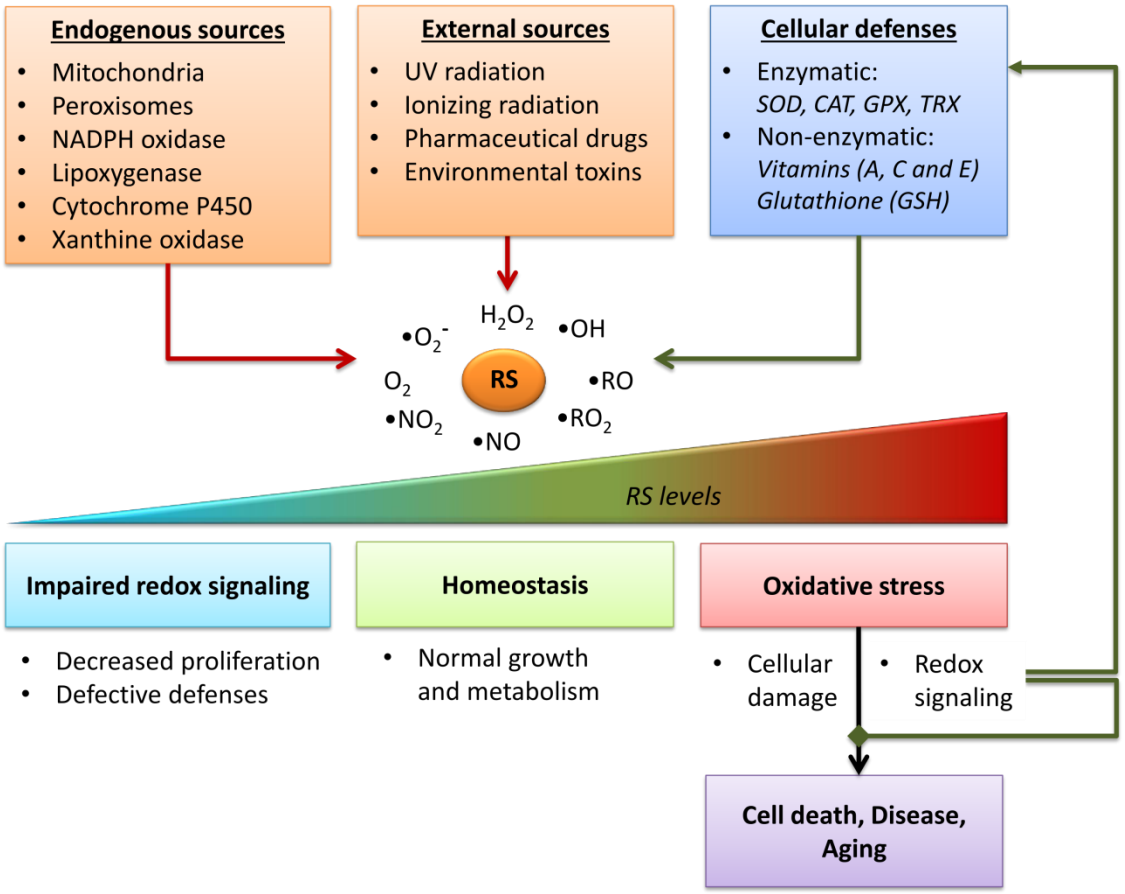


Figure 1.2).

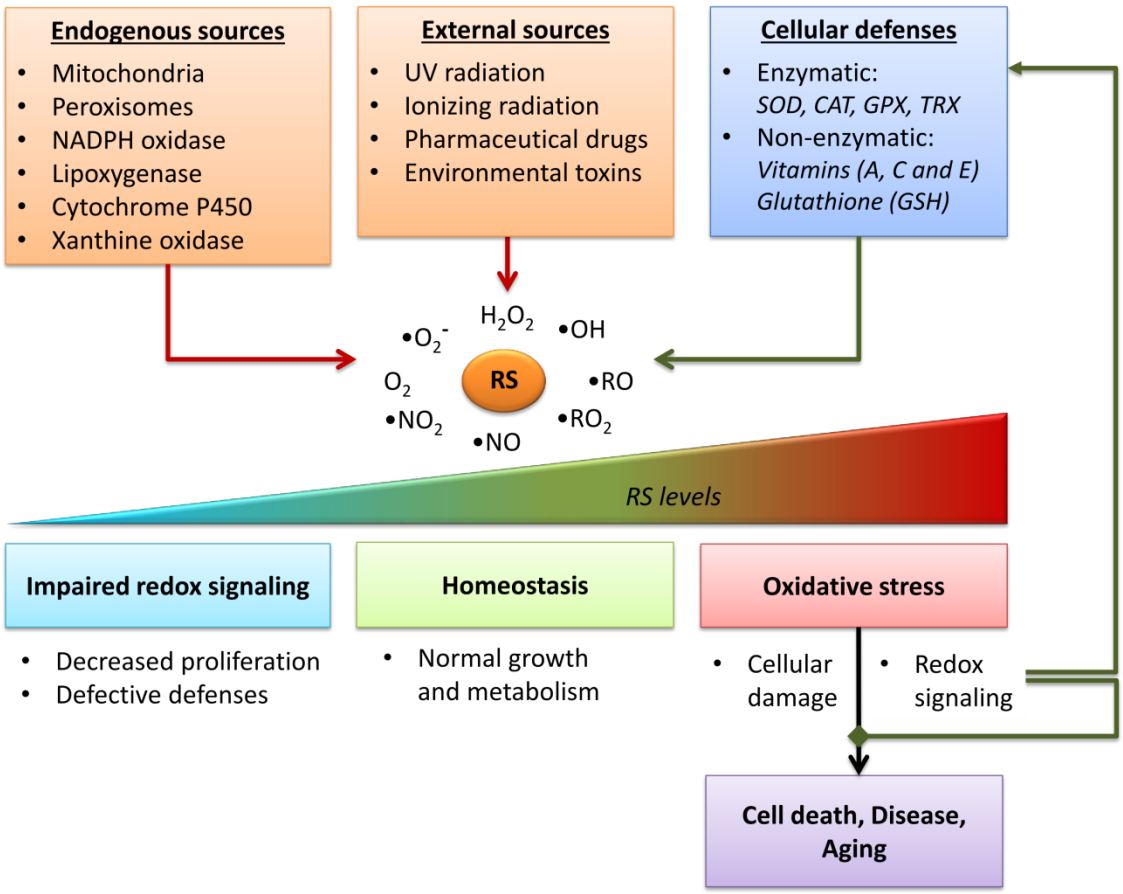


Figure 1.2 Sources and consequences of oxidative stress

Schematic shows common sources of RS (red arrows), antioxidant mechanisms (green arrows), typical RS and their effects on the cell. Endogenous sources of RS include mitochondria (primary source of RS), peroxisomes and various enzyme systems, while exogenous sources include pollutants/toxins, xenobiotics and ionizing/UV radiation. RS are counterbalanced by enzymatic antioxidant systems and non-enzymatic antioxidants, keeping the cellular homeostasis. Oxidative stress arises when defensive mechanics are overwhelmed by the RS, leading to cellular damage and induction of redox signalling (such as Keap1-Nrf2 pathway) which upregulates defensive mechanisms. Prolonged stress leads to cell death and pathology and is associated with ageing. On the other end of the spectrum, aberrantly low RS levels impede redox signalling and certain cellular functions (such as proliferation or immune system). Adapted from (50).

Notably, while RS are often generalized, it is important to take into account that different RS have very different chemistries. For example, the hydroxyl radical (OH•) reacts with all biological macromolecules at diffusion rate (half-life approximately 10^{-9} sec) and is thus highly reactive and unlikely to be intercepted by antioxidants, but also highly localized in its activity. On the contrast, the peroxyl radical (ROO•) has a half-life in seconds and is therefore much less reactive (51). Interestingly, while not considered RS, transition metals, i.e. iron (Fe) and to a lesser extent copper (Cu), play a critical role in the biochemistry of oxidative stress by mediating Fenton chemistry reactions (conversion of H_2O_2 into OH•) (51). Therefore, control

over free metal content in the cell, e.g. by iron chelating proteins such as ferritin and transferrin, is a critical part of cellular defences against oxidative stress (52). A brief summary of biochemistry of oxidative stress given here, and the reader is referred to (5,6) for further detail.

Oxidative damage to nucleic acids

Reactive chemical species damage nucleic acids by induction of single and double stranded breaks, base modifications and cross-linkage with proteins (5). This leads to genetic damage via point mutations and DNA deletions. While single-stranded DNA breaks are well-tolerated by the cell, double-stranded breaks potentially lead to cell death, and mutations in coding DNA or promoter regions impede cell functions (49). Notably, physiological levels of superoxide anion (O2-•), hydrogen peroxide (H_2O_2) and nitric oxide (NO•) have not been associated with direct DNA damage, and the majority of RS-induced genotoxicity is likely caused by hydroxyl radical (OH•), including hydroxyl produced from H_2O_2 in presence of DNA associated iron and copper (i.e. via Fenton reactions) (53). Genotoxic effects of RS have been linked to various diseases such as cancer and accelerated ageing (54).

Lipid peroxidation

Lipid damage by RS is caused by lipid peroxidation, chain reaction initiated by free radicals such as hydroxyl radical (OH•), superoxide anion (O_2^-•) or peroxyl radical (ROO•) in the presence of oxygen and metal ions. Reaction of free radical and fatty acid produces lipid peroxide which in turn can react with another lipid (55). As this reaction chain can cause significant lipid disruption even with low levels of free RS molecules, it poses a considerable threat to the cell. Lipid disruption due to peroxidation leads to changes in lipid bilayers, alters membrane permeability and has the potential to inactivate membrane bound receptors and enzymes (54,56). Lipid peroxidation products have been used as markers of oxidative stress in degenerative conditions such as chronic obstructive pulmonary disease (54) and neurodegeneration (57).

Protein damage by reactive oxidative species

Proteins are subject to oxidative damage by oxidation of amino acid residues, potentially causing a change of conformation (for example by breakage of disulfide bonds) or loss of activity by altering critical residues. Free radicals can also cause protein fragmentation and alter protein conformation, cause changes in hydrophobicity and protein aggregation (5,54). These effects lead to loss of protein function and precipitation (58). Protein damage can affect

numerous biochemical pathways in the cell; for example, it can impair the function of redox sensors (such as Keap1) or transcription factors (such as Nrf2), causing dysregulation of protein biosynthesis and response to oxidative stress. Oxidative damage to DNA repair mechanisms can decrease these activities and aggravate the genotoxic effects of RS. Damage to metal chelating proteins can release free metals and enhance RS toxicity due to increased OH• production (53). Products of protein oxidation have been associated with autoimmune diseases, atherosclerosis, neurodegenerative diseases and ageing (53,57).

1.2.4 Radical-free oxidative stress and redox code

Oxidative stress is commonly associated with free radicals and direct damage to cellular macromolecules, e.g. lipid peroxidation and DNA damage (5). Recent research, however, supports the notion that oxidative stress is, at least in part, caused by non-radical species such as H_2O_2, and does not necessarily involve damage to cellular components (59). For example, it has been shown that elevated H_2O_2 levels can lead to cell death by inducing apoptosis via cellular signalling systems without causing direct oxidative damage to the cell (60). The "redox hypothesis" and recently formulated "redox code" model state that: 1) Redox elements, e.g. cysteine residues, play a critical role in cellular signalling; 2) Activity of these elements is regulated by common "control nodes", e.g. GSH and thioredoxin; 3) Redox elements form kinetically and spatially separated circuits; and 4) These form the cellular redox signalling network. According to this model, oxidative stress is the product of disruption of "redox code", the cellular signalling network, and cannot be explained only by an imbalance between antioxidants and pro-oxidants. (59,61)

According to the "redox hypothesis" model, oxidative stress has two distinct mechanisms (Figure 1.3). Free radicals (e.g. superoxide and hydroxyl radicals) are mainly responsible for direct damage to cellular components (e.g. lipid peroxidation, DNA breakage and protein oxidation), while non-radical species (e.g. H_2O_2) cause indirect oxidative stress through disruption of cellular signalling (e.g. by oxidation of critical thiols of transcription factors or related proteins leading to pathological gene expression or apoptosis). In addition, free radicals are likely to be a minor part of the RS produced by cellular metabolism, as exemplified by experiments with xanthine oxidase (59), suggesting that non-radicals form the majority of chemical species responsible for oxidative stress. Therefore, if non-radical species (e.g. H_2O_2) are the major causes of oxidative stress related pathologies via indirect damage, i.e. disruption to cellular signalling, the pathology can correlate to markers of macromolecular damage (e.g. lipid peroxidation) and be relatively insensitive to free radical scavengers (59,61).

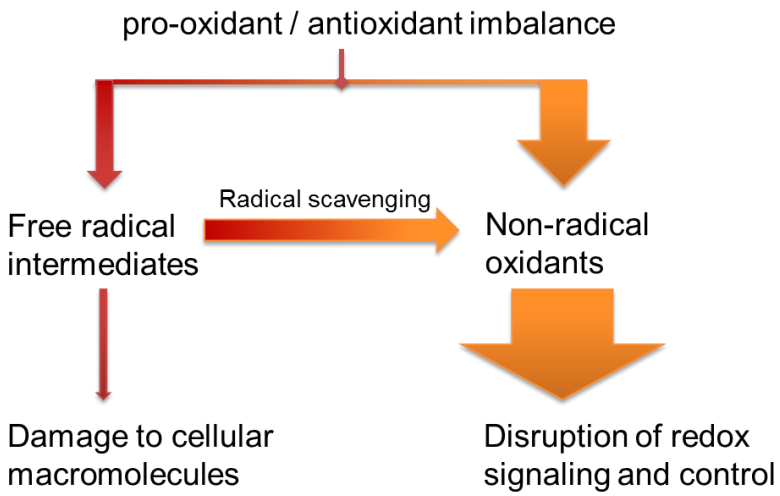


Figure 1.3 Redox hypothesis model, two mechanisms of oxidative stress

Pro-oxidant / antioxidant imbalance leads to production of free radical (e.g. $\bullet O_2^-$) and non-radical oxidants (e.g. H_2O_2), and the majority of RS are of non-radical type. Free radicals, responsible for direct damage to cellular macromolecules (left side), are further converted into non-radical species (e.g. $\bullet O_2^-$ to H_2O_2 by superoxide dismutase); therefore, chemical species produced during oxidative stress conditions are mainly non-radical (right) and generate oxidative stress by disruption of cellular signalling (e.g. leading into apoptosis). Schematic is based on (59,61).

1.2.5 Responses to oxidative stress

The Earth's oxidizing environment is by its very nature damaging to biological systems composed of oxidation-prone macromolecules such as proteins, lipid membranes and nucleic acids (5,6). Thus the change from anaerobic to aerobic environment on Earth, roughly 2.5 billion years ago (62), led to a selective pressure to evolve protection against the toxic effects of oxygen and other reactive (oxidative) species. As a result, current forms of life, with the exception of anaerobic bacteria and certain obligatory anaerobic fungi, flagellates and amoebae (63) possess a number of adaptations to oxidative stress. These can be roughly divided into mechanisms for prevention of oxidative stress, scavenging/interception of RS, and repair of oxidative damage (5).

Prevention of oxidative stress

The first line of defence includes prevention of oxidative stress by avoiding generation of RS from exogenous sources such as UV radiation and endogenous sources such as oxidative phosphorylation on mitochondrial membranes. UV avoidance and resistance are common adaptations, examples including movement away from sunlight and production of UV-absorbing pigments such as melatonin in mammals and mycosporine-like amino acids in algae

and some fungi (52,64). Organisms living in high environments exposed to high levels of solar UVR have also evolved photo-dependent DNA repair mechanisms (65).

Free metal ions, e.g. iron and copper, increase reactivity of RS by catalysing Fenton reactions of hydrogen peroxide to hydroxyl radical. Thus, metal chelation by enzymes such as ferritin and transferrin play a critical part in prevention of oxidative stress (66). Another strategy of prevention of oxidative damage is "prevention by diversion", i.e. by providing alternative targets for RS to minimize the damage to critical cellular components. The packaging of DNA into chromatin, where nucleic acid is shielded by histone proteins, is considered to be a form of this type of protection against oxidative attack; on the organism level, examples include tissues such as intestinal mucosal cells which serve as a protective layer that is rapidly replaced (52).

Scavenging and interception of RS

In order to cope with RS production and the threat of oxidative stress, cells have evolved an array of enzymatic and non-enzymatic mechanisms for scavenging and interception of RS, i.e. conversion of reactive species such as superoxide radical ($O_2^{-}\bullet$) into chemically inert species, for example water. Common scavenging mechanisms are illustrated in Figure 1.4.

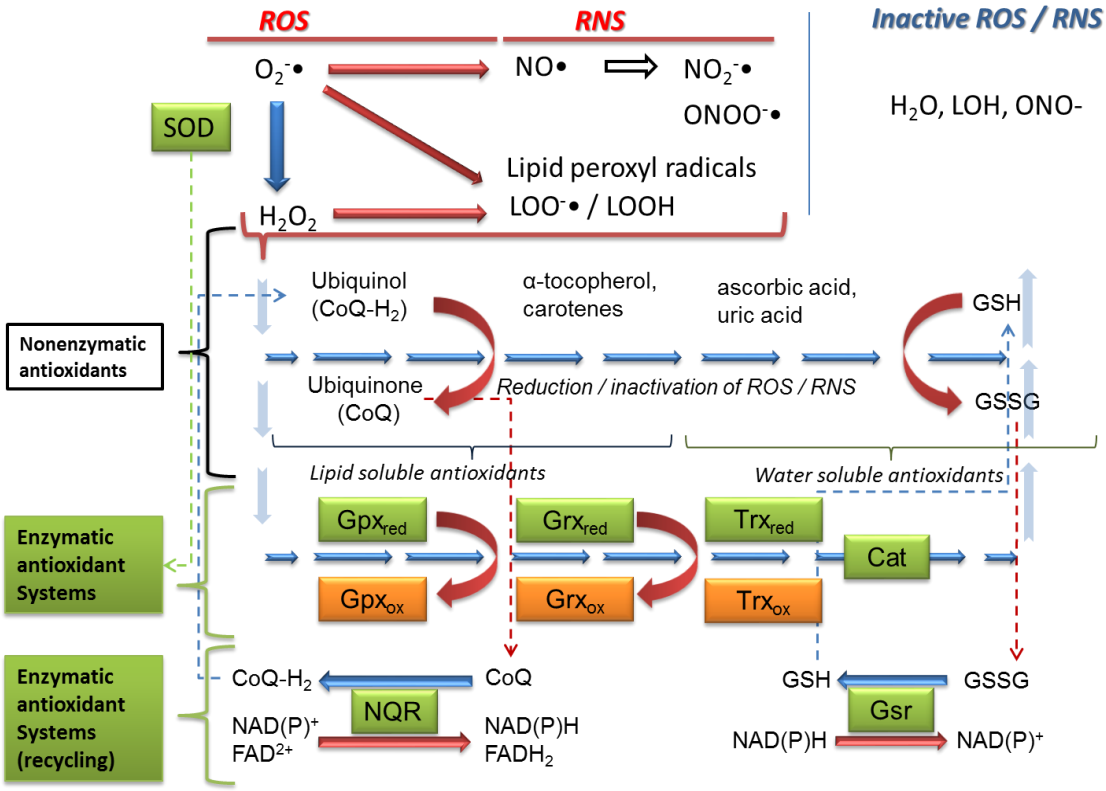


Figure 1.4 Antioxidant mechanisms

Various exogenous and endogenous mechanisms generate RS (ROS and RNS). Enzymatic and non-enzymatic antioxidants reduce these reactive chemical species and are in turn oxidized. Major small molecule antioxidants include uric and ascorbic acids, ubiquinol (CoQ) and glutathione (GSH). Major enzymatic systems include superoxide dismutase (SOD), catalase (Cat), thioredoxin (Trx), glutaredoxin (Grx) and glutathione peroxidase (Gpx); recycling enzymatic systems such as glutathione reductase (Gsr) and NAD(P)H quinone (oxido)reductase (NQR) turn oxidized antioxidants back into their reduced, functional forms. Red arrows show routes of oxidation, while blue arrows show reduction reactions. Figure is based on (67).

Antioxidants can be classified into non-enzymatic / small molecule antioxidants (examples are listed in Table 1.1) and macromolecular / enzymatic antioxidants (Table 1.2); Cellular localization and compound solubility also play important roles as lipid soluble compounds such as α-tocopherol protect biological membranes, while water-soluble compounds such as ascorbic acid are active in the cytosol (5).

Table 1.1 Major non-enzymatic antioxidants

Table lists examples of non-enzymatic antioxidants. Lipid soluble antioxidants protect cell membranes against RS, while water soluble antioxidants scavenge RS in the cytosol. The list is compiled from (5,6).

Antioxidant scavenger	Other names	Solubility
Ubiquinol	CoQ (CoQ_{10}, CoQ-H_2)	Lipids
α-tocopherol	Vitamin E	Lipids
β-carotene		Lipids
Retinol	Vitamin A	Lipids
Uric acid		Water
Glutathione	GSH	Water
Ascorbic acid	Vitamin C	Water

Enzymatic antioxidants include enzymes that catalyse the multi-step conversion of RS to water, as well as enzymatic systems for recycling of non-enzymatic and enzymatic antioxidants (67). Examples are shown in Table 1.2.

Table 1.2 Major antioxidant enzymes

Table lists some of major antioxidant enzyme families, common acronyms and catalysed reactions. Most of those enzymes exist in many variants with similar structure and many are also capable of catalysing reactions not listed in the table. *: It should also be noted that acronyms and names of these enzymes are not consistent through the literature. The list is compiled from (54,68).

Antioxidant scavenger	Acronym*	Reaction
Superoxide dismutase	SOD	$O_2^-\bullet \rightarrow O_2$; $O_2^-\bullet + 2H^+ \rightarrow H_2O_2$
Superoxide reductase	SOR	$O_2^-\bullet + 2H^+ \rightarrow H_2O_2$
Catalase	Cat	$2\ H_2O_2 \rightarrow 2\ H_2O + O_2$
Glutathione reductase	Gsr, GTRx	$GSSG + NADPH + H^+ \rightarrow 2\ GSH + NADP^+$
Glutathione peroxidase	Gpx, GPx	$ROOH + 2GSH \rightarrow ROH + GSSG + H_2O$
Thioredoxin	Trx, TRX	Reduces oxidized protein-S_2 / Prx_{ox}
Thioredoxin reductase	TR, TrxR	$Trx_{ox} + NAD(P)H + H^+ \rightarrow Trx_{red} + NAD(P)^+$
Glutaredoxin	Grx	Reduces oxidized protein-S_2 / Prx_{ox}
NADPH quinone (oxido)reductase	NQR	CoQ → CoQ-H_2
Peroxiredoxin	Prx	Reduces H_2O_2, R-O_2H

1.2.6 Pathology of oxidative stress

Numerous diseases have been associated with oxidative stress, including neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis (69); cardiovascular diseases, e.g. atherosclerosis, coronary artery disease and hypertension (70); diabetes and related complications such as vascular and kidney damage (71); and various types of cancer caused by RS-induced mutagenesis such as hepatocellular carcinoma and lung cancer (72).

In addition, oxidative stress is also regarded as one of the causes of age-related degeneration (4). These disorders tend to co-occur with increase in markers of oxidative stress in tissue samples of afflicted patients, as compared to healthy tissues. For example, most of the diseases are associated with an increase in the products of lipid peroxidation by free radicals, a depletion of reduced glutathione, an increase in the levels of oxidized nucleotides and markers of oxidative damage to proteins (57). In addition, experiments with *in-vivo* mouse models with deficient responses to oxidative stress, e.g. due to Nrf2 knockout, show increased sensitivity to neurodegenerative diseases (73), damage by external stressors such as cigarette smoke (74), and increased carcinogenesis when exposed to genotoxic agents (75). Although the presence of markers of oxidative damage in a large number of pathologies is well-supported, it should be noted that co-occurrence of oxidative damage and disease does not necessarily imply causality, i.e. markers of oxidative damage in disease tissue does not by itself prove that oxidative stress causes the pathology; this subject is discussed further in 1.2.8.

1.2.7 Oxidative stress and ageing

Among the definitions of ageing, a commonly used one states that ageing is "a collection of time dependent, universal, intrinsic, progressive and deleterious changes in cells and tissues of an organism, eventually leading to degeneration and death" (76). While the phenomenon of ageing has been of considerable interest to people since the dawn of human history, with written sources describing ageing dating to early antiquity (77), modern study of ageing dates to the late 19th century (77). Although dozens (by some sources hundreds (76)) of theories of ageing have been postulated, a prominent theory directly linked to redox biology research is the free radical theory of ageing (FRTA) (76,78,79). Influenced by the discovery of free radicals in biological tissues in 1954 (2), Harman proposed the FRTA in 1958, hypothesizing that ageing is caused by accumulation of cellular damage inflicted by free radicals (4). The theory evolved over time and its current incarnation is the mitochondrial free radical theory of ageing (MFRTA). Based on the fact that oxidative phosphorylation on the mitochondrial membranes is a major source of RS in aerobic metabolism (80), MFRTA states that RS produced by the

respiratory chain causes damage to the mitochondria; in turn, this causes cellular degeneration followed by tissue damage that leads to the degenerative changes associated with ageing (81). According to this theory, ageing can be considered a degenerative disorder caused by oxidative stress and is potentially treatable by prevention or alleviation of the stress.

1.2.8 Controversies and difficulties in redox biology

Since the discovery of free radicals in biological systems, of oxidative damage to cellular components, and of enzymatic antioxidant systems, research in the field is now referred to as redox biology and has been largely driven by the notion that free radicals are deleterious and cause a number of diseases and disorders unless properly countered by antioxidants. Thus, oxidative stress has been connected to a variety of diseases, as well as cancer and ageing, and major efforts have been undertaken to identify and classify antioxidants for, presumably beneficial, dietary supplementation.

While these views were never without critics, recent years have seen a slow shift of paradigm away from the classical "RS are damaging, antioxidants are beneficial" view (1,4) to the "redox code" hypothesis which emphasizes the balance of redox signalling as a central theme (13,61). In addition, a number of commonly used methods for assessing antioxidants and oxidative stress have been put into doubt, resulting in a call for greater standardization of the vocabulary, models and methods used in research of oxidants and antioxidants (17,82).

In light of these ongoing changes, here follows a short review of controversies of redox biology, focusing on concepts, e.g. oxidative stress and antioxidants, as well as recent results obtained from clinical trials of antioxidant therapies. This review is not indented to be comprehensive, to criticize individual research, or to answer these controversies; it is instead intended to draw attention to, and to illustrate the difficulties of redox biology, which will undoubtedly be avenues for future research.

1.2.8.1 What is oxidative stress?

The term oxidative stress, originally introduced by Sies as *an imbalance between oxidants and antioxidants* is one of the basic concepts of redox biology. Yet it is an unfortunate term as it implies stress caused by oxidizing chemical species but does not postulate a clear hypothesis. Indeed, in his recent review, Sies himself stated that an *oxidative stress hypothesis* has not been formulated up to now and he discouraged the unselective use of this term (83).

Recent research identified that the RS in the cell are not always harmful and that removal of RS might be deleterious to the organism. For example, moderate increase in RS levels during

aerobic exercise has been shown to be beneficial in mouse models (84), while direct antioxidants have been found to inhibit human killer cell activity (10) and prevent the health-promoting effects of physical exercise in humans (15). Furthermore, as discussed in section 1.2.8.4, clinical trials of antioxidant therapy have failed to demonstrate benefits to human health, and instead suggested that certain antioxidants are deleterious to humans (85). The role of RS in neurodegeneration (86) and in ageing has also been put into doubt (87,88). Altogether, these studies suggested that the original definition of oxidative stress might be outdated.

Should the term "oxidative stress" be abolished? Several suggestions have been made in recent years. For example, it might be appropriate to abolish the term altogether and instead focus on specific processes and (bio)chemistry (83). Another option would be to replace the term "oxidative stress" with the "redox code" hypothesis that serves the same role, but is more in line with recent models of redox biology that attribute the RS-caused pathology to the dysregulation of redox signalling rather than by oxidative burden (61). Another suggestion has been to classify the oxidative stress into "stress levels" according to biological effects, e.g. mild, physiological stress that triggers antioxidant defences as opposed to high level, toxic, stress that damages the cell (89). Finally, there has been a proposal to split oxidative stress into positive stress, "oxidative eustress", that upregulates cellular defences, e.g. by Nrf2 activation, and "oxidative distress" that inflicts cellular damage by disruption of cellular signalling and oxidation of cellular components (13).

Notably, none of these suggestions currently has a significant following in the scientific community, as evident from the vast majority (above 99.9%) of recently published research articles that used the term "oxidative stress" [data is based on web of science term analysis of articles published in the 2007-2017 period, results are not displayed]. Thus, whilst arguably outdated, the term "oxidative stress" seems to be entrenched and is likely to stay in use, at least in the near future.

1.2.8.2 What is an *in-vivo* antioxidant?

Similar to oxidative stress, antioxidant is a very vague term and recently published guidelines discourage its use (17), yet just like the term oxidative stress, it seems unlikely to be abolished in the near future. Halliwell and Gutteridge defined an antioxidant as "a substance that, when present at low concentrations compared with those of an oxidisable substrate, significantly delays or prevents oxidation of that substrate" (26). The problem with this term is that most chemical species can be labelled as antioxidants by choice of appropriate chemical assay (26).

Nonetheless, defining antioxidant *in vitro* is reasonably straightforward, as long as definition includes the assay or test; for example, the definition in the food industry could be "a substance which slows down the natural oxidation and decay of food". Thus, the term "antioxidant" is clearly measurable and has a significant utility in an *in vitro* environment.

Compounds with *in vitro* antioxidant activity, however, are not necessarily active *in vivo.* That is because the oxidation in biological systems is a result of multiple chemical species and the chemistry of RS is in no way uniform, e.g. half-lives of peroxyl radical (ROO•) and hydroxyl radical differ by ~10 orders of magnitude (51). In addition, biological systems possess a number of highly active "antioxidant" systems including enzymatic systems and endogenously produced small-molecule antioxidants, e.g. glutathione, α-tocopherol and ascorbic acid (52,53,90). Therefore, compounds with activity lower than endogenous *in-vivo* antioxidants are unlikely to significantly reduce oxidative stress (16,53). Recent studies of antioxidant therapies failed to demonstrate health benefit in humans (91), suggesting that tested direct antioxidants such as ascorbic acid and α-tocopherol do not have significant *in-vivo* activity, when introduced in the diet.

By contrast, "indirect antioxidants" - compounds that activate the biosynthesis of endogenous RS neutralizing enzymes - are known to protect cells against oxidative stress *in vivo.* For example, indirect antioxidant sulforaphane (SFN) protects mouse fibroblasts from UV-A induced oxidative stress (92), and resveratrol protects human lung epithelial cells from cigarette smoke (93). These compounds are not chemical antioxidants and protect cells by induction of the Nrf2 pathway. Despite being called "antioxidants", these compounds are, in fact, oxidizing chemical species and activate redox signalling in the cell by *oxidising* cysteine residues of sensor proteins such as Keap1 (94–96). This implies that RS, in general, are "indirect antioxidants", which is contrary to the very definition of an antioxidant. In addition, it is also unclear whether compounds that *prevent* oxidative damage by other mechanisms, e.g. metal chelation or blocking of UV radiation, should be termed indirect antioxidants as antioxidants do prevent RS mediated damage, but do not conform to the above definition.

1.2.8.3 Does oxidative stress cause disease and/or ageing?

As discussed in 1.2.5, biomarkers of oxidative cell damage, such as the products of lipid peroxidation, have been found in a large number of diseases, including neurodegenerative diseases and various cancerous tumours (57). Co-occurrence of oxidative damage, however does not provide proof that oxidative stress *causes* a disease, because confounding factors can cause both the oxidative damage and the disease (86).

Similar criticism can be raised for genetically-modified animal models: co-occurrence of a certain mutation or phenotype and markers of oxidative stress does not prove causality. For example, genetic mutation might be introducing a phenotype that causes the tissue damage (e.g. by dysregulation of cell signalling leading to apoptosis or necrosis), and the tissue damage causes oxidative stress; thus, the observed effects would be symptoms of disease and markers of oxidative stress, but the main cause of disease is not the oxidative stress (53).

Therefore, unless the biochemical pathways to a disease are known (which is not the case for most diseases), a proof that oxidative stress *causes* the disease would require an experiment where introduced oxidative stress can be correlated to increased incidence of the disorder compared to the control. Alternative proof, if less strict, is to provide strong evidence that reduction of oxidative stress, e.g. by antioxidant therapy, alleviates the disease. These arguments were raised by Halliwell and Gutteridge as early as 1984 (8) and their subsequent work added a large amount of evidence to back their argument (16,53,97–99). A strong argument *against* causal link between oxidative stress and (most) diseases is that cell death *causes* oxidative stress. The ruptured cell releases metal ions and hydrogen peroxide into surrounding tissue and induces an immune response, all of which elevates RS concentration. As tissue damage is an integral part of any degenerative disease and tissue damage causes oxidative stress, markers of oxidative damage will thus always be present in disease-affected tissues, even if tissue damage is caused by stress-independent pathology, e.g. dysregulation of metabolic pathways leading to necrotic cell death (86,99,100). The schematic of this model is shown in Figure 1.5.

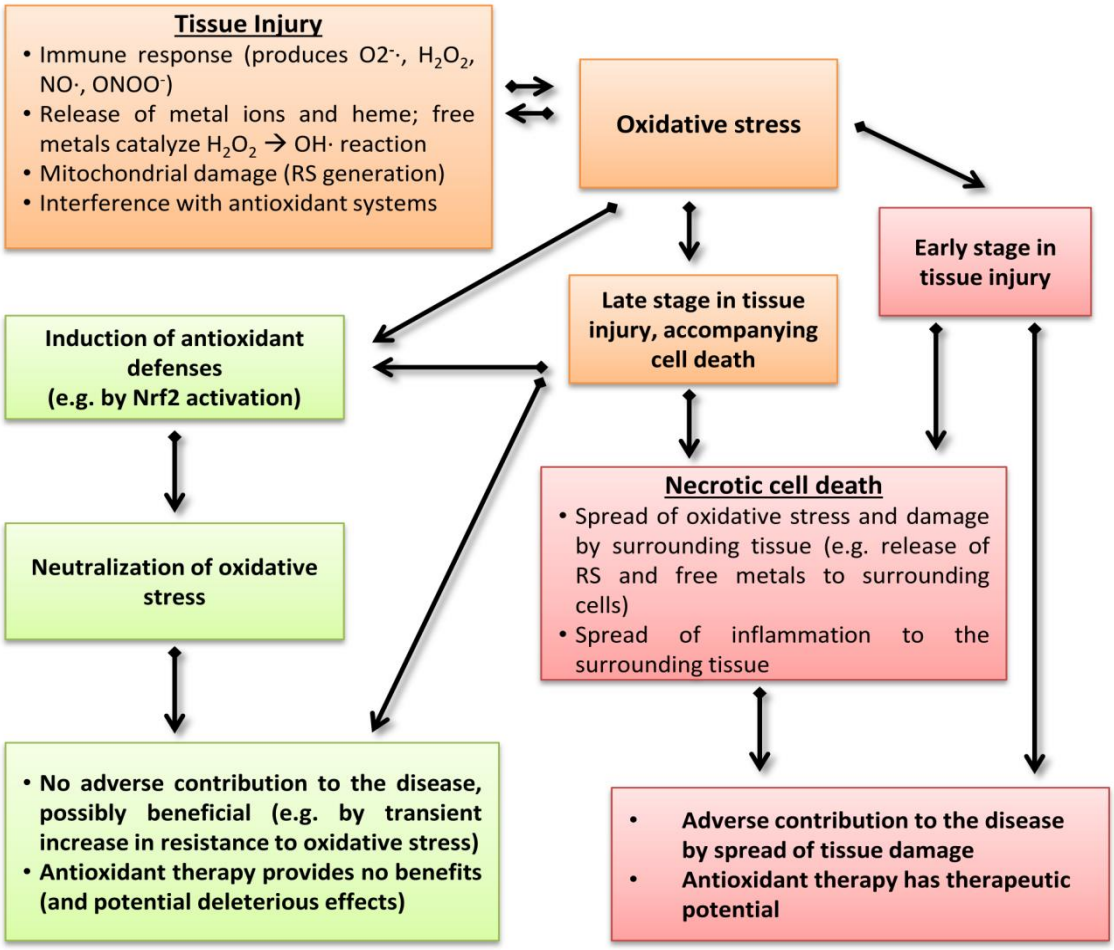


Figure 1.5 Oxidative stress and tissue damage

Schematic illustrates relations between tissue injury, oxidative stress and diseases. Processes leading to the oxidative stress are coloured orange; mechanisms by which oxidative stress causes a disease are coloured in red, and mechanisms by which stress alleviates the pathology are denoted in green. Schematic is based on (16,53,99).

1.2.8.4 Failure of antioxidant based therapies

Multiple studies based on addition of direct antioxidants such as ascorbic acid and α-tocopherol to cell culture or to mouse diet demonstrated that antioxidant therapy has a potential to prevent cancer (101–104). However, when translated to double-blinded clinical trials with human subjects, these therapies failed to demonstrate a benefit in treatment of degenerative diseases or in cancer prevention. For example, a recently concluded trial of Coenzyme Q10 for alleviation of early Parkinson's disease (105) demonstrated no benefit; a combination of antioxidants for treatment of Alzheimer's disease fared no better, failing to influence progression of the disease (106). Multiple studies have been conducted to demonstrate the cancer-preventive effects of antioxidants including β-carotene, vitamin E, vitamin C, selenium, retinol, zinc, riboflavin, and molybdenum, and none were shown to provide significant protection against cancer (107).

Finally, recent meta-analyses of 78 trials involving a total of ~300,000 participants determined that there is no evidence to support the use of antioxidant supplements for prevention or treatment of human diseases (85,91). In addition, authors of these meta-studies concluded that increased intake of vitamin E, beta-carotene and vitamin A increased the mortality rate of study participants (85,91). While the causes of the observed increase in mortality rate have not been determined in these studies, it is possible that increase intake of chemical, direct, antioxidants reduces the endogenous, Nrf2-mediated, cytoprotection leading to increased susceptibility to oxidative stress. Antioxidant therapy has been previously found to inhibit the beneficial effects of aerobic exercise in men, presumably by inhibiting endogenous activation of cytoprotection (15).

While these meta-studies and failures of clinical trials provide strong evidence against antioxidant therapies for treatment of human diseases, there is a number of arguments to take into account: 1) Dietary antioxidants rarely cause much change in biomarkers of oxidative damage in humans (98,108); 2) The majority of trials used a synthetic form of vitamin E and used much lower doses compared to cell-based and mouse model studies (109,110); 3) these trials did not investigate indirect antioxidants such as SFN which were demonstrated to have an effect *in vivo* (111); 4) the chemistry of RS is in no way uniform and scavenging of RS by non-enzymatic systems is unlikely to be a main mechanisms of prevention of RS-induced oxidative stress *in vivo* (51,52,90); and 5) multiple antioxidant-based drugs have successfully passed clinical trials and are currently in use, e.g. Edaravone for the treatment of ischemic shock and n-aceylcysteine for paracetamol overdose (108).

In summary, while oxidative stress has been implicated in a large number of diseases, antioxidant therapies have so far been unsuccessful in alleviating these disorders in humans. These failures are likely due to the overly optimistic approach based on the free radical scavenging by "direct antioxidants" (110). While this approach is attractive due to its simplicity, it is unlikely that all diseases associated with biomarkers of oxidative damage are in fact caused by oxidative stress (86). Thus, it is not unexpected that the "one-size fits all" approach has failed to provide a universal treatment to the vast array of biochemically very different diseases and disorders in humans. A better understanding of biological processes involved in these disorders and the recently postulated "redox code hypothesis" (see 1.2.4) might be of considerable utility in explaining mechanisms of diseases associated with oxidative stress. Finally, a more tailored approach, perhaps by targeting cellular response networks such as the Keap1-Nrf2 pathway (see 1.3) by "indirect antioxidants" may offer the potential to succeed where general antioxidants have failed (16,53).

1.3 Nrf2-mediated response to oxidative stress

Cells are constantly challenged by electrophiles endogenously generated as by-products of oxidative metabolism, and by the UVR, environmental oxidants and pollutants. In animal cells, one of the master regulators of redox metabolism is transcription factor Nrf2 which lies at the centre of an intricate signalling system composed of the transcription factor Nrf2, its repressor Keap1, the transcription enhancer ARE and a number of other components. This system was discovered at the turn of the 20th century and has since attracted considerable attention in redox biology and medicine (112,113).

1.3.1 Discovery of Nrf2/ARE/Keap1 pathway

A common promotor of stress response genes was first described by Picket et al. in 1990 as the *β-NF responsive element*, an upstream enhancer of the rat glutathione-S-transferase (GST) gene, which responds to beta-naphthoflavone (β-NF) (114). The DNA sequence of the β-NF responsive element was analysed by mutagenesis and selective deletions and the sequence was identified as 5'-RTGACNNNGC-3' (where R denotes A or G, and N denotes any nucleotide). The same study recognized that this enhancer responded to H_2O_2 and phenolic antioxidants that undergo redox cycling, prompting the renaming of the β-NF responsive element to the antioxidant response element (ARE) (115). Whilst the acronym ARE is widely used (and will be used in throughout this thesis), the electrophile response element (EpRE) has also been suggested as a more appropriate name as most ARE inducers are, in fact, electrophiles or function by generating H_2O_2 (116).

Bioinformatics analysis of putative ARE sequences by Nerland (2007) (117) and mutagenesis experiments by Hayes et al. (2003) identified a number of differences in ARE sequences of mouse, rat and human genes and the authors argued that one universal ARE sequence is unlikely (118). The study suggests rather that AREs show significant variation between the species, even for closely related organisms such as mouse and rat (118). Notably, the evolutionary history and relationships of AREs are currently unknown and worthy of future investigation.

1.3.1.1 Nrf2, discovery and function

Nuclear factor erythroid-2-related factor 2 (Nrf2) was first described by Moi et al. in 1994 as the transcription activator of β-globin protein expression in human immortalised myelogenous leukemia cell-line K562. The study also identified that this protein was expressed *in-vivo* in mice and in nonerythroid cell lines HeLa and Raji (119). Based on its homology to *Drosophila*

melanogaster Cap-n-Collar (CNC) proteins, the Nrf2 was classified as a member of the "Cap-n-Collar" (CNC) subfamily of basic leucine zipper (bZIP) transcription factors (119). In 1995, a study with chicken erythroid cells, progenitor cells of megakaryocytes and erythrocytes, determined that the chicken homolog of the mouse CNC transcription factor p45 NF-E2 (mouse Nrf2), likely plays a role in avian erythropoiesis. This protein was named erythroid cell-derived protein with CNC homology (ECH) (120). Interestingly, studies of Nrf2 knockout mice ($Nrf2^{-/-}$ genotype) found no visible phenotype, leading to the conclusion that Nrf2 gene function is not necessary for mouse development, growth, or fertility (121). However, this study was based on the assumption that Nrf2 plays a role in hematopoiesis in mice and effects of external factors such as oxidative stress, UV radiation or mutagens were not tested on these Nrf2 knockout animals.

The Nrf2 function as a promotor of cellular defences was postulated based upon the similarity of the ARE (5'-*TGAC*NNNGC-3') to Nrf2 binding site (5'-RTGASTCAGCA-3'; R denotes A or G, while S denotes C or G) and prompted another venture into Nrf2 knockout mice experiments by Yamamoto et al. (122). The study, published in 1997, was based on induction of phase-2 detoxifying enzymes glutathione-S-transferase (GST) and NAD(P)H:quinone oxidoreductase (NQO1) by the phenolic antioxidant butylated hydroxyanisole (BHA), previously found to induce activity of antioxidant enzymes in mice tissues (122). Results demonstrated that BHA does not upregulate GST or NQO1 production in Nrf2 knockout mice, as opposed to heterozygous ($Nrf2^{-/+}$) and wild type ($Nrf2^{+/+}$) mice where these enzymes were upregulated when exposed to BHA. Notably, this experiment also demonstrated that other CNC proteins, such as Nrf1 and Nrf3, do not share Nrf2's function. Furthermore, the study used an electrophoretic mobility shift assay (EMSA) to demonstrate that Nrf2-ARE binding required heterodimerization of Nrf2 with a small Maf (sMaf) protein. While the study used a MafK protein, the authors suggested that other sMaf proteins might also heterodimerize with Nrf2 (122), which was later confirmed from studies of sMaf knockout mice, where sMafs were found to be essential for embryonal development and for upregulation of ARE-controlled genes (123). Since then, research based on gene knockout mice models (122,124–126) and high-throughput *–omics* experiments such as microarrays, proteomics and ChIP-seq (125,127–130) have identified Nrf2 as a regulator of biosynthesis of over two hundred genes. Many of these genes are involved in a response to oxidative stress, drug transport, phase-2 detoxification or drug metabolism, strongly implying that Nrf2 is a master regulator of mammalian stress response networks.

1.3.1.2 Discovery of inhibitors of Nrf2

Initial studies of the regulation of Nrf2 were performed by Yamamoto et al. (1999). Functional characterization of Nrf2 by sequence similarity comparison uncovered 6 domains conserved between the human Nrf2 protein and chicken homolog. These domains were named Neh1 – Neh6, where Neh stands for Nrf2-ECH homology (131). Of these, domain Neh1 was identified as a DNA-binding domain based on homology to the DNA-binding domains of other CNC-bZip proteins (131). The Neh2 domain was found to be critical for inhibition of Nrf2 activity, as demonstrated by enhanced ARE binding of Nrf2 protein when the Neh2 domain was deleted, and prompted search for Neh2 binding inhibitors of Nrf2 (131). Subsequent yeast 2-hybrid studies with Neh2 as bait identified that Neh2 binds to yeast protein with similarity to a previously uncharacterized human protein designated KIAA0132. This protein was named the Kelch-like ECH-associated protein 1 (Keap1) (131). Further research characterized the structure of Keap1, while immunoprecipitation and cellular localization studies with fluorescent constructs (Nrf2-GFP and Keap1-GFP) verified its role as the inhibitor of Nrf2. These studies concluded that Keap1 binds to Nrf2, through the Neh2 domain, and sequesters Nrf2 in the cytoplasm, thereby preventing migration of Nrf2 to the nucleus and Nrf2-DNA binding. Two years later, it was demonstrated that electrophiles such as diethylmaleate (DEM) and catechol cause release of Nrf2 from Keap1 (131,132). Finally, in 2003 and 2004, Yamamoto's group performed studies of Keap1-Nrf2 and identified that the mechanism of Nrf2 inhibition also includes degradation of Nrf2. Keap1 was recognized as mediator of the ubiquitination/degradation process, functioning as an adaptor for the cullin-3 RING ubiquitin ligase (CRL) complex (133,134). Further research identified that the Neh2 domain of Nrf2 contains high affinity ETGE and low affinity DLG motifs that bind the Keap1 dimer (135). This led to a proposal called the "hinge and latch" model of Keap1-Nrf2 interaction; according to this model, Keap1-Nrf2 binding is a two-step process, where ETGE binding creates an "open" Keap1-Nrf2 conformation and is followed by DLG binding resulting in a "closed" complex required for CRL-mediated ubiquitination (135). A current model of Keap1-Nrf2 interaction, called the "cyclic sequential attachment and regeneration model of Keap1-mediated degradation of Nrf2", was suggested based on live-cell imaging studies (136) and explains the interaction as a four-step process, with the Nrf2 ubiquitination step and Nrf2 release with the subsequent degradation step following the two-step "hinge and latch" model (136,137).

In addition to Keap1-based regulation of Nrf2, a number of other mechanisms have been suggested to control Nrf2 activity, indicating that this transcription factor is under much more complex regulation than originally postulated. As of 2016, proteins including chromo-

ATPase/helicase DNA-binding protein (CHD) (138), cAMP response element-binding protein (CREB)-binding protein (CBP) (139), β-transducin repeat-containing protein (β-TrCP) (140) and retinoid X receptor (RXR)α (141) have all been suggested to play a role in Nrf2 regulation, alongside other mechanisms such as micro-RNAs (142), phosphorylation (143,144), Notch signalling (145), endoplasmic reticulum stress (146) and unfolded protein stress (147). These mechanisms are briefly reviewed below.

1.3.2 Components and regulation of Nrf2-Keap1-ARE pathway

Discovery of Keap1-based inhibition of transcription factor Nrf2 led to the naming of the Nrf2/ARE/Keap1 pathway. Later research, however, discovered that a multitude of other components are involved in the regulation of Nrf2. Here follows the brief review of the current model describing other components involved in Nrf2 regulation, with emphasis on canonical (i.e. Keap1 dependent) regulation of Nrf2 activity.

1.3.2.1 Nuclear factor erythroid-2-related factor 2 (Nrf2)

Nrf2 is the major regulator of mammalian redox signalling, implicated in the control of a large number of genes involved in various cellular functions, mainly associated with cellular defences and detoxification (see 1.3.3).

Structurally, Nrf2 is a modular protein consisting of Nrf2-ECH homology (Neh) domains 1 to 7 (Figure 1.6). The Neh1 domain of Nrf2 forms the "cap-n-collar" basic leucine zipper (CNC-bZip) structural motif and is responsible for dimerization with small Maf proteins and DNA binding (128). The Neh2 is responsible for canonical, Keap1-mediated, negative regulation of Nrf2 activity. Neh2 is composed of DLG and ETGE motifs responsible for recruitment of Keap1, which leads to Cullin (Cul)3–RING (really interesting new gene)-box protein (Rbx)1 ubiquitin ligase complex (CRL^{Keap1}) mediated ubiquitination and degradation of Nrf2 (131). The C-terminal of the Neh3 domain recruits the chromo-ATPase/helicase DNA-binding protein (CHD)6, and upregulates Nrf2 activity (138). Both Neh4 and Neh5 are transactivation domains responsible for binding of CBP and/or RAC (139). The Neh6 domain binds the dimeric β-transducin repeat-containing protein (β-TrCP) via the DSGIS and DSAPGS motifs. Neh6 is responsible for an alternative pathway for ubiquitination and degradation of Nrf2, mediated by the S-phase kinase-associated protein 1 (Skp1)–Cul1–Rbx1 core E3 complex (SCFb-TrCP) (140). Finally, the Neh7 domain downregulates Nrf2 activity by recruiting retinoid X receptor (RXR)α (141).

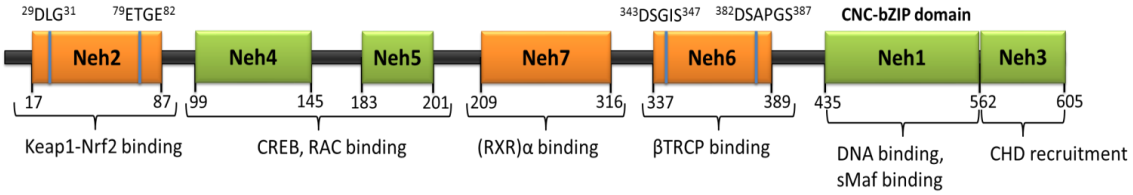


Figure 1.6 Structure of Nuclear factor erythroid-2-related factor 2 (Nrf2).

Schematic displays functional domains and conserved motif, with amino acid numbering based on human Nrf2 protein. Domains involved in activation of Nrf2 are coloured green, while domains involved in its inhibition are coloured orange. Nrf2 binds to its inhibitor Keap1 via Neh2 motifs DLG and ETGE; domains Neh4 and Neh5 represent transactivation domains that recruit receptor-associated coactivator (RAC)3 and/or cAMP response element-binding protein (CREB)-binding protein (CBP); domain Neh7 inhibits Nrf2 activity by physical association with the retinoid X receptor (RXR)α; Neh6 domain is involved in Nrf2 degradation mediated by dimeric β-transducin repeat-containing protein β-TrCP; Neh1 is DNA binding and small Maf co-factor binding domain and C-terminal Neh3 domain is transactivation domain involved in recruitment of chromo-ATPase/helicase DNA-binding protein (CHD). Schematic is based on (96).

1.3.2.2 Kelch-like ECH-associated protein 1 (Keap1)

The Keap1 (Figure 1.7) protein is a Broad complex, Tramtrack and Bric-a-Brac (BTB)-Kelch protein that forms a homodimer under normal cell conditions (148). Keap1 contains a BTB domain responsible for Keap1:Keap1 dimerization and for binding of Cullin protein (148), and an intervening (IVR) domain and six Kelch/double glycine (DG) repeats. These form a six bladed beta-propeller tertiary structure (149) which binds to the Neh2 domain of Nrf2 (131). Keap1 contains a number of "electrophile sensing" cysteine residues that modulate its Nrf2 inhibition activity (i.e. electrophile attack on Cys151, Cys226/Cys613, Cys-273/Cys-288 and Cys-434 residues causes conformational changes of Keap1 and prevents degradation of Nrf2 by the CRL^{Keap1} complex) (96).

The view that Keap1 is the primary inhibitor of Nrf2 activity has been questioned and multiple "non-canonical" Nrf2 regulation mechanisms, described briefly in 1.3.2.4, have also been identified (150). The role of Keap1 is firmly supported, however, because disruption of Keap1 is sufficient for a major increase in cellular Nrf2 activity (151,152). In addition, somatic mutations of Keap1 characteristic for certain cancers (153,154) result in upregulation of Nrf2, as does inactivation of Keap1 by micro RNAs (142) or hypermethylation of the Keap1 promotor (155).

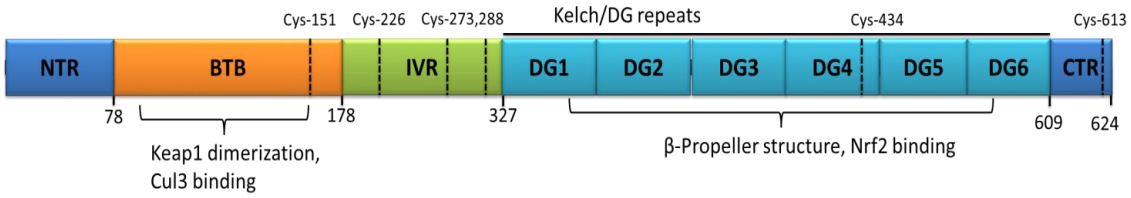


Figure 1.7 Structure of Keap1

Figure displays the schematic of Keap1 protein. The Broad complex, Tramtrack and Bric-a-Brac (BTB) domain, coloured orange, is responsible for formation of Keap1 dimer and Nrf2 binding; Kelch-repeat domain, also referred to as double glycin (DG) repeat domain, forms six-bladed β-propeller structure and binds DLG and ETGE motifs of Neh2 domain of Nrf2. Intervening (IVR) domain is comprised of residues between BTB and Kelch repeats. Cysteine residues that function as electrophile sensors are denoted above the schematic.

1.3.2.3 Keap1 based regulation of Nrf2 activity

While the cell is not under oxidative stress, Nrf2 is subject to constitutive proteasomal degradation in a step-wise CRL^{Keap1}-dependent manner; accordingly, Nrf2 has a short half-life and very low basal activity in unstressed cells (134). Specifically, the Neh2 domain of Nrf2 contains an ETGE motif which binds a β-propeller subunit of the Keap1 dimer with high affinity, and a DLG motif with low Keap1-binding affinity. Nrf2 binding and degradation is a four step process (see Figure 1.8), with Nrf2-ETGE binding to a β-propeller subunit of the Keap1 dimer to form an "open" conformation, followed by Nrf2-DLG binding to the other Keap1 β-propeller subunit to "close" the Keap1-Nrf2 complex. This complex enables recruitment of Cul3-Rbx1 and the ubiquitination of Nrf2. Ubiquitinated Nrf2 is released from the Keap1 dimer and degraded in the proteasome, freeing the Keap1 dimer to bind another molecule of Nrf2 (136,137). This model of Keap1-mediated inhibition of Nrf2 is supported by live cell imaging experiments (136,137) and is an extension of previously postulated "hinge and latch" two-step model Keap1-Nrf2 interaction via the ETGE and DLG domains of Nrf2 (135).

Oxidative stress, i.e. increased concentration of RS in the cell, leads to conformational changes of Keap1 which block the degradation cycle by preventing ubiquitination of Nrf2 and subsequent release from Keap1; this in turn results in saturation of free Keap1 in the cell and allows for nuclear translocation of newly synthesized Nrf2. Specifically, Keap1 contains a number of "RS-sensing" cysteine residues (Cys151, Cys226, Cys273, Cys288, Cys434 and Cys613) which are prone to modification by electrophiles (e.g. cysteine -SH group oxidation that leads to formation of disulphide bridges between the Keap1 cysteine residues). These changes do not prevent Keap1-Nrf2 binding, but disrupt the binding of the "closed" Keap1-Nrf2 complex to the CRL ubiquitination complex, thus stopping the degradation of Nrf2.

Interestingly, recent studies suggest that Keap1 is not reactivated after conformational changes by electrophiles and is instead degraded (136). Electrophile activators of Nrf2, such as tert-butylhydroquinone (tBHQ), sulforaphane (SFN), Bardoxolone, dimethyl fumarate (DMF) and the majority of other known Nrf2 inducers, also function by this mechanism by targeting critical Cys residues of Keap1 (156,157). Thus, these compounds, while sometimes referred to as "indirect antioxidants" or "Nrf2-activating antioxidants", are in fact, pro-oxidants. These compounds oxidize biological macromolecules and have chemical properties similar to RS rather than conventional direct antioxidants such as α-tocopherol or ascorbic acid.

Non-electrophile activators of Nrf2, e.g. small-peptides developed by Wells et al. (158,159) and recently developed small-molecule Nrf2 activators (160), function by competitive inhibition of Keap1-Nrf2 binding. These compounds inhibit the complexation of Keap1-Nrf2 by binding to the Keap1 β-propeller structure responsible for DLG and ETGE binding. While less studied than electrophile-based Nrf2 activation, targeting of the Keap1-Nrf2 binding site by small compounds or peptides has potential to upregulate Nrf2 without the danger of "off-target" effects, as these compounds and peptides do not depend on electrophile modification of cysteine residues, which is non-selective and likely to affect a multitude of cysteine-containing proteins of other cellular signalling networks (157).

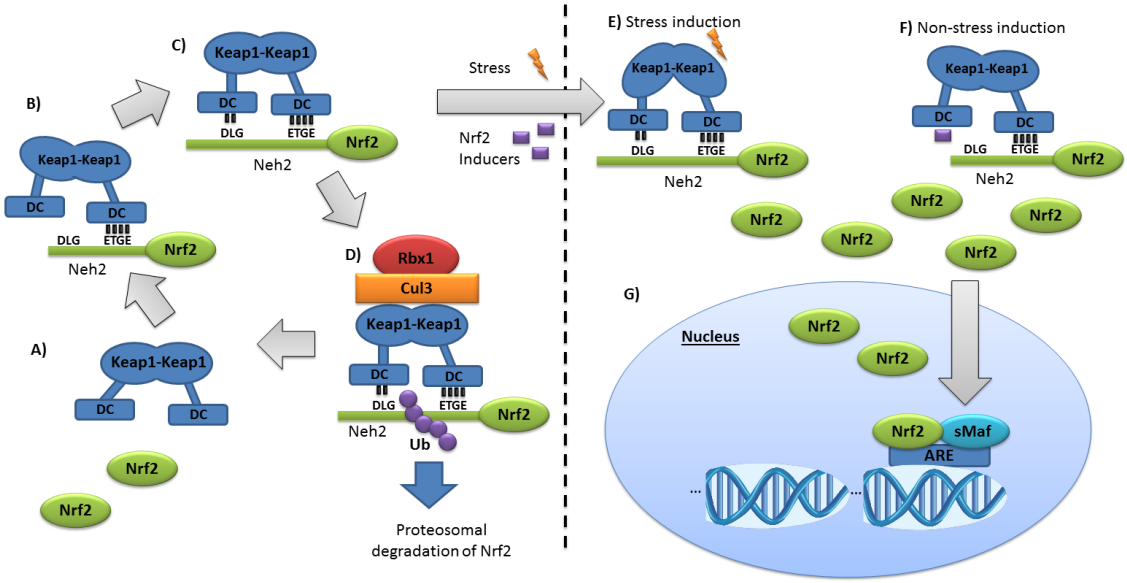


Figure 1.8 Model of Keap1 mediated regulation of Nrf2

Figure displays "cyclic sequential attachment and regeneration model of Keap1-mediated degradation of Nrf2" (A-D) and a model of Nrf2 induction by stress and non-stress inducers (E-G). A single unit of Keap1-Keap dimer binds ETGE motif of Nrf2 Neh2 domain with high affinity, forming "open" conformation (B); this is followed by low-affinity binding of second Keap1 unit to Nrf2 Neh2 DLG motif and results in "closed" conformation; (C) the "closed" conformation interacts with Cul3 and Rbx1 to form protein complex which ubiquitinates Nrf2 (D); ubiquitinated Nrf2 is released from Keap1 and degraded, freeing Keap1 dimer to bind next Nrf2 molecule (A). Stress (e.g. presence of electrophiles such as H_2O_2) causes the oxidation of critical Cys residues of Keap1 which leads to change in conformation, preventing the ubiquitination of Nrf2 by Cul3 (E); this leads to saturation of free Keap1 in the cytosol, allowing the newly synthesized Nrf2 to translocate to the nucleus (G). Alternatively, non-electrophile induces of Nrf2 can bind to one of DC beta propellers of Keap1-dimer to prevent the "closing" of the Keap1-Nrf2 complex (F) which again leads to conformation unsuitable for binding of Cul3 to Keap1, saturation of Keap1 and translocation of Nrf2 to the nucleus (G). Keap1 dimer is shown in blue, with DC (DLG repeat and C-terminal domain structure) representing β-propeller structure of Keap1 responsible for Keap1-Nrf2 binding; Nrf2 is shown in green and Rbx1 and Cul3 proteins are shown in red and orange. Schematic is based on (135,136,157).

1.3.2.4 Non-canonical regulation of Nrf2 activity

The Keap1-Nrf2 system has been studied in great detail over the last decade and a large number of Keap1-independent mechanisms of Nrf2 regulation have been suggested. These mechanisms are described in-depth in recent general reviews of Nrf2 regulation (96,156,161) as well as specialized reviews of Nrf2/Notch signalling (162), Nrf2/NF-κB signalling (163), epigenetic regulation of Nrf2 (155), the role of β-TrCP (164) and the role of nuclear receptors such as RXRα (165). The work presented here is primarily based on the "canonical", Keap1-mediated regulation of Nrf2. Thus, the "non-canonical" mechanisms of Nrf2 control are not considered critical for the results presented in Chapters 2 to 5 and are reviewed here only briefly.

Keap1 independent degradation of Nrf2

While Keap1-mediated degradation of Nrf2 is considered to be the primary mechanism of Nrf2 degradation, Nrf2 is also degraded at a slow rate in stressed cells, implying the existence of Keap1-independent mechanisms. Mutagenesis analysis of Nrf2 identified that the Neh6 domain is necessary for Nrf2 degradation under conditions of oxidative stress (166). Further studies showed that the Neh6 domain of Nrf2 contains DSGIS and DSAPGS motifs that bind the β-TrCP protein (140,167). β-TrCP binds to Nrf2 via the β-propeller WD40 domain, forming the ubiquitination complex with SCF E3 ubiquitin ligase. This complex ubiquitinates Nrf2, resulting in proteosomal degradation of Nrf2 (140). Notably, a number of kinases, e.g. GSK-3 and 'priming' kinases required for GSK-3 activity, have also been implicated in β-TrCP mediated degradation of Nrf2, indicating that phosphorylation plays a role in regulation of this process (167,168).

In addition to Keap1 and β-TrCP, Hrd1 is an integral endoplasmic reticulum (ER) membrane E3 ubiquitin ligase, which has been found to downregulate Nrf2 levels during liver cirrhosis, as identified in tissue samples from liver cirrhosis patients and confirmed by *in-vivo* Hrd1 knockout mouse models (146). This downregulation of Nrf2 is probably caused by ER stress, i.e. stress resulting from unfolded proteins in ER but has so far only been reported in liver cirrhosis (156).

Transcriptional and epigenetic regulation of Nrf2

The gene encoding Nrf2is designated the NFE2L2 gene and is considered constitutively expressed, but is under a number of regulatory mechanisms, as evident from differences in expression levels between tissues (169). For example, Nrf2 levels are considerably higher in mouse liver and kidney compared to muscles, lungs and brain (169). Analysis of the mouse NFE2L2 promotor identified a xenobiotic response element (XRE) and two sequences similar to XRE; these are enhancer sequences involved in upregulation of xenobiotic response genes, e.g. the CYP super family enzymes involved in phase I detoxification (170). In addition, the promotor region also includes two ARE elements, an NF-κB binding site and an a12-O-tetradecanoylphorbol-13-acetate-response element that allows the transcriptional activation of NFE2L2 by oncogene KrasG12D via transcription factors c-Jun and c-Fos. Therefore, NFE2L2 transcription is potentially regulated by Nrf2 itself, as well as xenobiotics, inflammation processes, and oncogenes (156). Furthermore, it has been shown that Nrf2 levels might be influenced by other stress signalling pathways, for example response to fasting (171).

In addition to promotor-level regulation, evidence exists for epigenetic regulation of *NFE2L2* transcription by DNA methylation and histone modification (155). For example, Nrf2 is repressed epigenetically during neuronal development in mice, resulting in negligent expression of an Nrf2-dependent antioxidant response in mature mouse neurons (172). Another example is hypermethylation of the NFE2L2 promotor that increases Nrf2 activity in tumorigenic cells, but not in non-tumorigenic cells in a transgenic adenocarcinoma of mouse prostate (TRAMP) model (173).

Finally, microRNAs have been implicated in both positive and negative regulation of Nrf2 levels (142). An example of negative regulation of Nrf2 is homozygous sickle cell anemia disease, where erythrocytes of patients with the disease have reduced tolerance to oxidative stress due to Nrf2 downregulation by the microRNA designated miR-144 (174). In contrast, Nrf2 is upregulated by microRNAs in some types of breast cancer by miR-200a mediated degradation of Keap1 (175).

Regulation of Keap1-Nrf2 binding by competitive inhibition with endogenous proteins

Human and mouse cells contain a number of proteins with affinity for the Nrf2 Neh2 domain and Keap1 beta-propeller structure. For example, the autophagy cargo receptor and signalling adaptor protein p62/SQSTM1 contains a STGE binding motif similar to the Neh2 motif ETGE responsible for binding Keap1. Keap1-p62 binding was demonstrated *in vitro* and in mouse cell lines, where an increase of p62 led to increased activity of Nrf2 (176). Similarly, the stress response cyclin-dependent kinase inhibitor $p21^{Cip1/WAF1}$ interacts with Neh2 ETGE and DLG motifs and competes with Keap1 binding to prevent ubiquitination and degradation of Nrf2, leading to upregulation of Nrf2 controlled genes such as NQO1 and HO-1. This was demonstrated by a combination of cell-based assays and p21 knockout mouse models (177). A number of other ETGE motif-containing proteins, e.g. peptidase DPP3, kinase IKKβ, tumour suppressor gene BRCA associated protein PALB2 and phosphoglycerate mutase PGAM5, were reported also to interact with Keap1, but the importance of these interactions has yet to be determined *in vivo* (96).

1.3.3 Role of Nrf2 in cellular functions and disease

At the cellular level, Nrf2 has been associated with the upregulation of a large number of genes implicated in a variety of processes including detoxification (phase I, II and III), biosynthesis of antioxidants such as glutathione (GSH) and thioredoxin (TRX), lipid metabolism, carbohydrate metabolism, iron metabolism (96) and mitochondrial function (178). Examples of genes under Nrf2 control are listed in Table 1.3. The commonly quoted number for Nrf2 controlled genes is "hundreds" (157), presumably based on ChIP-sequencing and microarray studies (125,129). Malhotra et al. (2010) used ChIP-sequencing and microarrays to identify 244 genes under Nrf2 control in mouse embryonic fibroblasts (125), while Chorlie et al. (2012) detected 242 Nrf2 bound genomic regions in human lymphoblastoid cells (129).It should be noted, however, that other studies have reported numbers of genes under Nrf2 transcriptional regulation ranging from as high as 700 (128), to as low as 23 (179). For example, Chip-Seq experiments by Hirotsu et al. (2012) found 702 genomic regions co-occupied by Nrf2 and MafG (but of these, only 66 genes were upregulated by pro-oxidant diethyl maleate in a Hepa1 cell line, as detected by microarrays (128)); Chorlie et al. (2012) identified 242 Nrf2 bound genomic regions, yet the authors noted that the expression of about two-thirds of the candidate genes were likely to be Nrf2-dependent, putting the final number closer to ~150 (129). Yet another study based on quantitative proteomics of Nrf2 knockout mouse livers, found widely different numbers of Nrf2 regulated proteins (38 and 108 in two repeats of the experiment) (180).

Table 1.3 Examples of genes positively regulated by Nrf2 in mice and humans

Table lists examples of Nrf2 upregulated genes associated with various cellular functions. Genes supported by multiple independent studies are marked by *. Table is based on (96) and references therein.

General function	Symbol	Description
Detoxication	ADH7*	alcohol dehydrogenase class 4 mu/sigma chain
Phase I	AKR1B1*	aldo-keto reductase family 1, member B1 (and 1B8 and 1B10)
(Drug oxidation,	AKR1C1*	aldo-keto reductase family 1, member C1 (and 1C2 and 1C3)
reduction and	ALDH1A1*	aldehyde dehydrogenase 1 family, member A1
hydrolysis)	ALDH3A1*	aldehyde dehydrogenase 3 family, member A1 (and 3A2)
	ALDH7A1	aldehyde dehydrogenase 7 family, member A1
	CBR1	carbonyl reductase 1 (and 3)
	CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1
	CYP2B9	cytochrome P450, family 2, subfamily B, polypeptide 9 (and 10)
	EPHX1*	epoxide hydrolase 1, microsomal
	PTGR1*	prostaglandin reductase 1 (also called LTB4DH)
	NQO1*	NAD(P)H:quinone oxidoreductase 1
Detoxication	GSTA1*	glutathione S-transferase class Alpha 1 (and A2, A3 and A4)
Phase II	GSTM1*	glutathione S-transferase class Mu 1 (and M2, M3, M4, M6 and M7)
Drug conjugation	GSTP1*	glutathione S-transferase class Pi 1 (and P2)
	MGST1*	microsomal glutathione S-transferase 1 (and 2)
	SULT1A1	sulfotransferase family, cytosolic, 1A, member 1 (2)
	UGT1A1	UDP glucuronosyltransferase 1 family, polypeptide A1 (and 1A6)
	UGT2B7	UDP glucuronosyltransferase 2 family, polypeptide B7 (and 2B34)
Detoxication	ABCB6*	ATP-binding cassette, subfamily B (MDR/TAP), member 6
Phase III	ABCC1*	ATP-binding cassette, subfamily C (CFTR/MRP) member 1
(Drug transport)	ABCC2*	ATP-binding cassette, subfamily C (CFTR/MRP) member 2
	ABCC3*	ATP-binding cassette, subfamily C (CFTR/MRP) member 3
	ABCC4*	ATP-binding cassette, subfamily C (CFTR/MRP) member 4
	ABCC5	ATP-binding cassette, subfamily C (CFTR/MRP) member 5
Antioxidant,	GCLC*	glutamate-cysteine ligase, catalytic subunit
GSH-based	GCLM*	glutamate-cysteine ligase, modifier subunit
	GGT1	gamma-glutamyltransferase 1
	GLRX	glutaredoxin 1
	GLS	glutaminase
	GPX2	glutathione peroxidase 2
	GPX4	glutathione peroxidase 4
	GSR1*	glutathione reductase
	SLC6A9	glycine transporter
	SLC7A11	cystine/glutamate transporter
Antioxidant,	PRDX1*	peroxiredoxin 1
TXN-based	PRDX6*	peroxiredoxin 6

	SRXN1*	sulfiredoxin-1
	TXN1*	thioredoxin
	TXNRD1*	thioredoxin reductase 1
Carbohydrate metabolism and NADPH regeneration	G6PD*	glucose-6-phosphate 1-dehydrogenase
	HDK1	hexokinase domain containing 1
	IDH1*	isocitrate dehydrogenase 1
	ME1*	malic enzyme 1, NADP+-dependent, cytosolic
	PGD*	6-phosphogluconate dehydrogenase
	TALDO1*	transaldolase
	TKT	transketolase isoform 1
	UGDH	UDP-glucose dehydrogenase
Iron metabolism	BLVRA	biliverdin reductase A
	BLVRB	biliverdin reductase B [flavin reductase (NADPH)]
	FECH*	ferrochelatase
	FTH1*	ferritin, heavy polypeptide 1
	FTHL12*	ferritin, heavy polypeptide-like 12
	FTHL17	ferritin, heavy polypeptide-like 17
	FTL1*	ferritin, light polypeptide
	HMOX1	heme oxygenase (decycling) 1
Transcription regulation	AHR	arylhydrocarbon receptor (AhR)
	CEBPB*	CCAAT/enhancer-binding protein (C/EBP), β
	MAFG*	MafG protein
	NFE2L2	nuclear factor-erythroid 2-like 2
	PPARG*	peroxisome proliferator-activated receptor gamma (PPARγ)
	PPARGC1 B	peroxisome proliferator-activated receptor gamma coactivator 1-beta
	RXRA*	retinoid X receptor alpha (RXRα, or NR2B1)
	KEAP1	Kelch-like ECH-associated protein 1
Lipid metabolism	ACOT7	acetyl-CoA thioesterase 7
	ACOT8	acetyl-CoA thioesterase 8
	ACOX1	acetyl-CoA oxidase 1
	ACOX2	acetyl-CoA oxidase 2, branched chain
	AWAT1	acetyl-CoA wax alcohol acyltransferase 1
	CES1G	carboxylesterase 1G
	CES1H	carboxylesterase 1H
	SCD2	stearoyl-CoA desaturase-2
	LIPH	lipase, member H
	PLA2G7	phospholipase A2, group vii (platelet-activating factor acetylhydrolase)
	PNPLA2	patatin-like phospholipase domain containing 2

1.3.3.1 Role of Nrf2 in drug metabolism and oxidative stress response

Xenobiotic metabolism is a set of biochemical pathways for modification, detoxification and efflux of compounds foreign to an organism's biochemistry, such as drugs and poisons. Often referred to as drug metabolism, these pathways are divided into three phases. Phase I includes oxidation and modification of xenobiotics by enzymes such as the cytochrome P450 oxidase (CYP) superfamily. Modified xenobiotics, which are often toxic, are conjugated to polar compounds, e.g. glutathione, in phase II reactions. These reactions are mediated by transferases such as glutathione S-transferases and UDP glucuronosyltransferases. Phase III drug metabolism includes further processing and efflux of xenobiotics by a variety of membrane transporters of the multidrug resistance protein (MRP) family (181,182).

Numerous studies, based on high-throughput technology such as Chip-seq and microarrays (125,129) as well as more traditional molecular biology approaches (122,124–126), have identified that numerous genes involved in drug metabolism phases I – III and antioxidant systems are under control of an ARE promotor and directly induced by upregulation of Nrf2 (see Table 1.3 for examples) (96). These roles were confirmed by multiple *in-vivo* mice experiments, with Nrf2-knockout ($Nrf2^{-/-}$) mice found to be more susceptible to oxidative stress-based diseases and drug-induced toxicity. Activation of Nrf2 by pharmacological (i.e. Nrf2 activating compounds such as sulforaphane) or genetic (i.e. knockdown of the Keap1 encoding gene) approaches alleviated the $Nrf2^{-/-}$ mouse susceptibility to oxidative stress. For example, the phenolic antioxidant butylated hydroxyanisole (BHA) upregulates a number of phase-2 detoxification genes, such as *GST*, *GCLC* and *NQ01*, in wild-type mice, while upregulation is lost in $Nrf2^{-/-}$ mice (122). Similarly, compared to wild type mice, Nrf2 knockout animals are sensitive to environmental pollutants such as diesel exhaust (183), carcinogens such as benzo[a]pyrene (184), and drug toxicity, e.g. acetaminophen (185). Whilst the majority of research in this area has been based on mouse models, human cell-line studies have confirmed animal model findings; for example, Nrf2 upregulation was found to protect human kidney cells against cytotoxicity of cisplatin (186), and to protect human epithelial cells against cigarette smoke (93).

1.3.3.2 Role of Nrf2 in lipid and carbohydrate metabolism

Multiple studies using microarray and proteomics analyses detected the upregulation of gene products involved in lipid and carbohydrate metabolism using cell cultures treated using Nrf2 activators such as sulforaphane (SFN) (127,156). These results were confirmed *in vivo* using mice treated with sulforaphane (SFN) (187). In addition, Nrf2 knockout mice were found to downregulate genes encoding products controlling lipid and glucose metabolism (188).

Furthermore, different research groups reported that Nrf2 knockout mice are prone to fatty liver when placed on methionine- and choline- deficient diets, implying that Nrf2 is important for proper metabolism of lipids (156).

A recent study by Brachs et al. (2016) (189) however found no evidence for Nrf2 regulation of glucose or lipid metabolism in adult mice, as measured in an *in-vivo* mouse model with a hepatic Keap1 knockdown performed using siRNAs. The authors explained these results by arguing that previously observed Nrf2 regulation of lipid and glucose metabolism was, in fact, the result of toxic effects from the use of Nrf2 activators such as CDDO-Im, CDDO-Me and Oltipraz (189). The study of Brachs et al. (2016) performed whole genome expression analysis using microarrays and targeted qPCR analysis of Nrf2, Keap1 and selected ARE regulated genes, and so experimental errors due to cross-hybridization and analogue nature of the microarray signal (190) were unlikely.

1.3.3.3 Role of Nrf2 in degenerative diseases

Since the discovery of the role of Nrf2 in protection against oxidative stress, the role of Nrf2 in disease has been subject to extensive study. The Nrf2 protein has been associated with protection from a number of diseases including neurodegenerative disorders (191), e.g. Alzheimer's disease and Parkinson's disease, various types of cancer (192), liver diseases (193) and kidney diseases (194). The role of Nrf2 in disease is reviewed in detail in (73,191,195), but briefly:

Neurodegenerative diseases such as Parkinson's disease (PD), Alzheimer disease (AD) and multiple sclerosis are associated with inflammation and oxidative stress, as evinced by markers of oxidative-stress induced damage found in affected tissues and blood plasma of patients (see (195), ref 3-31). A protective role of Nrf2 was studied in animal models of these diseases, and it was found that upregulation of Nrf2 alleviated the symptoms of these neurodegenerative conditions, while Nrf2 knockouts caused increased severity in symptoms (73).

For example, Nrf2 knockout mouse models of PD show greater loss of dopaminergic neurons and more severe lysis of astrocytes and microglia cells in the brain, as opposed to wild type mouse disease models (196). Nrf2 overexpression in astrocytes was demonstrated to protect against development of PD. Similar protective effects were observed in wild type mice, but not Nrf2 knockout models, when Nrf2 levels were upregulated by a Keap1 knockdown, or by treatment with Nrf2 activators such as SFN (197). As with PD, Nrf2 knockout mice models of multiple sclerosis showed early onset and faster progression of the disease, when compared to wild type mice. (198). Nrf2 upregulation, e.g. by dimethyl fumarate (199) and CDDO-TFEA

(200), moderated the symptoms and markers of oxidative stress in wild type mouse models of MS, whilst having no clear effect on Nrf2 knockout mice (199).

In addition to PD and MS, Nrf2 upregulation was found to afford protection in mouse models of other degenerative diseases, e.g. amyotrophic lateral sclerosis and Huntington's disease, as well as acute neurological disorders such as stroke and traumatic brain injury (73). Interestingly, the role of Nrf2 in AD is less certain; while human AD patients and mouse AD models exhibit a drop in Nrf2 levels (201,202), it is not clear if the decrease in Nrf2 activity caused the pathology in AD or are the result of AD-induced neuron death. In addition, different studies have reported conflicting results showing Nrf2 upregulation, implying that the role of Nrf2 in AD might be dependent on the stage of disease (73,195).

1.3.3.4 Role of Nrf2 in ageing

The free radical mitochondrial theory of ageing (FRMTA, see 1.2.7) postulates that ageing is caused by oxidative damage to the cell, which would indicate that Nrf2 might also play a role in the ageing process. A number of studies were recently conducted to test this hypothesis; for example, Kubben et al. (2016) (203) used a human fibroblast cell-line model for Hutchinson-Gilford Progeria Syndrome (HGPS), a premature ageing disorder, to test for activity of Nrf2 regulated genes. High-throughput screen of siRNAs identified significant repression of a large number of genes under ARE regulation in the HGPS cell line, while constitutive expression of Nrf2 prevented the formation of the HGPS phenotype (203). Another recent study by Pan et al. (2016) studied SIRT6, a protein associated with longevity, in human mesenchymal stem cells. This study demonstrated that knock-out of SIRT6 encoding gene dysregulated redox metabolism and caused premature ageing and sensitivity to oxidative stress in the stem cells (204). This study also identified that SIRT6 forms a protein complex with Nrf2, and this complex is presumably required for Nrf2 transcription regulation of genes such as *HO-1* (204).

1.3.4 The "Dark-side" of Nrf2

The foregoing discussions strongly suggest a role of Nrf2 in the control of cellular response to oxidative stress, and upregulation of Nrf2 as a potential therapy for oxidative stress related disorders such as cardiovascular and neurodegenerative diseases. Recent research, however, identified a correlation between Nrf2 upregulation and cancer proliferation and resistance to chemotherapeutic drugs, termed the "dark side of Nrf2". This has led to the recognition that constitutive expression of Nrf2 has potential to cause deleterious changes at both the cell and organism levels (112,205).

1.3.4.1 Effects of constitutive Nrf2 upregulation in mice

The first "dark side" effect of Nrf2 was observed during mice knockout experiments to generate Nrf2 and Keap1 deleted models. While Nrf2 deletion in mice was found to be of limited pathology, provided the mice were not exposed to excessive amounts of oxidative stress (121), Keap1 deletion was lethal and resulted in early postnatal mortality of mice due to hyperkeratosis of the oesophagus leading to problems with feeding (151). The pathology was caused by overexpression of Nrf2 rather than lack of Keap1 (151,206). A study of mice with a partial knockdown of Keap1 activity, generated using the Cre-loxP system for tissue specific gene knockdown (207), identified that long term overexpression of Nrf2 reduces the lifespan of mice (206). Interestingly, the same study also observed that an increase in Nrf2 levels conferred resistance to hepatic injury caused by acetoaminophen, but that the total inhibition of Keap1 in liver cells resulted in lower resistance compared to partial inhibition of Keap1 (206). In addition to identifying the importance of Nrf2 regulation in early mouse development, these studies demonstrated hormetic effects of Nrf2 upregulation in mice. The moderate increases in Nrf2 levels were beneficial, while higher and longer lasting upregulation had deleterious effects.

1.3.4.2 Role of Nrf2 in cancer development and resistance to chemotherapy

Nrf2 regulates transcription of numerous genes involved in cellular defences and has been associated with the prevention of oncogenesis (205,208), but the elevated Nrf2 levels have also been detected in various tumours (209–212) and associated with cancers resistant to chemotherapeutic drugs. This "dark side of Nrf2" was first identified in a study by Wang et al. (2008) who used human cell culture models to investigate Nrf2 levels in lung carcinoma, breast adenocarcinoma and neuroblastoma cancers. The authors determined that Nrf2 levels are increased in late-stage cancer cells, and that increase in concentration of Nrf2 enhances cell resistance to anti-cancer drugs. In addition, they demonstrated that Nrf2 knock-down by siRNAs sensitizes cells to chemotherapeutic agents (213).

Furthermore, extensive mapping of somatic mutations in human tumours identified a high number of mutations clustered in the Nrf2 Neh2 domain, in the proximity of the DNA sequence encoding ETGE and DLG protein motifs (214,215). These studies provided strong evidence that Nrf2 was constitutively expressed in certain types of tumours (213). Analysis of different types of cancers associated with loss of function of Keap1 demonstrated increased tumour aggressiveness and enhanced resistance to irradiation and chemotherapeutic agents such as cisplatin, carboplatin and etoposide (216). Hence, cancers where Nrf2 is upregulated would appear difficult to treat, leading to poor clinical prognosis (217,218), while cancers where Nrf2

activity is down regulated appeared less resistant to chemotherapy and thus more successful to treat (219). Based on these lines of research, cancers can be classified depending on Nrf2 activity levels, and measurement of Nrf2 levels has a potential to aid as a prognostic tool in cancer treatment (220).

The discovery of the "dark side" of Nrf2 led to the model of dual roles of Nrf2 in cancer: increased transcription of genes regulated by Nrf2 prevents cancer development, but enhances proliferation and drug resistance of existing cancer cells (220). Multiple studies (154,192,208) support this model. For example, Taguchi et al. (2010) reported reduced lifespan, but not increased cancer incidence, in Keap1 knockdown mice (206); Satoh et al (2013) found that Nrf2 knockout mice exhibit higher rate of carcinogenesis when exposed to the carcinogen urethane, but tumours in these animals are less aggressive and proliferate at slower rate compared to cancers in wild-type mice (221). The role of Nrf2 in cancer is currently an active area of research and is further reviewed in (112,154,192,211).

1.3.4.3 Role of Nrf2 upregulation in atherosclerosis and skin diseases

In addition to the "dark side" of Nrf2 in cancer, recent research has suggested that Nrf2 exhibits both beneficial and adverse effects in atherosclerosis, with diverse results observed based on the particular mouse model and the stage of the disease (222).

Increase in Nrf2 levels has been shown to protect against the early stage of atherosclerosis associated with oxidative damage. This was concluded from experiments with apolipoprotein E (ApoE) knockout mouse models and mouse models lacking the low density lipoprotein receptor. These animal models exhibited accelerated progression of early-stage atherosclerosis when transplanted with bone marrow cells of Nrf2 knockout mice (20,222). As discussed in 1.3.3.2, Nrf2 regulates the transcription of certain genes that influence lipid metabolism and the control of inflammation, and increase in Nrf2 levels might be responsible for acceleration of the inflammation associated later stages of atherosclerosis. This is because the Nrf2 knockdown *reduced* the symptoms of the disease in the mouse model of late stage atherosclerosis (223,224). Thus, the role of Nrf2 in atherosclerosis is still unresolved and provides another example that therapeutic increase in concentration of Nrf2 should be approached with caution.

The role of Nrf2 activation in skin diseases is currently controversial. While Nrf2 activators such as sulforaphane (SFN) and tert-butyl-hydroquinone (tBHQ) have been demonstrated to protect mouse and human skin *in vivo* (225,226) and in cell culture models (227), long term increase in concentration of Nrf2 was found to be deleterious in mouse animal models. For example, Schaefer et al. (2012) designed transgenic mice with constitutive expression of Nrf2

in keratinocytes and these mice developed severe skin pathology manifesting as epidermal thickening, hyperkeratosis and inflammation resembling lamellar ichthyosis (228). Their later research confirmed the effect for prolonged pharmacological activation of Nrf2 (by SFN and tBHQ) in wild type mice (229), suggesting that while short-term induction of Nrf2 protects the skin against oxidative damage, long term Nrf2 upregulation causes pathological alterations of the skin.

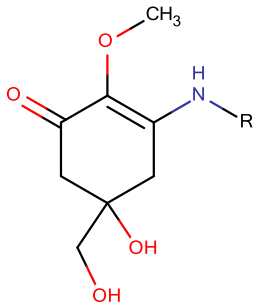
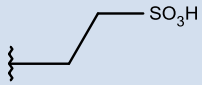
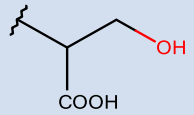
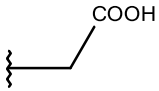
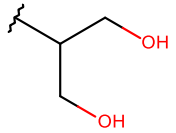
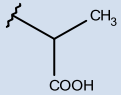
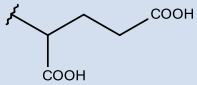
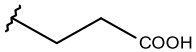
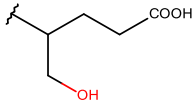
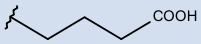
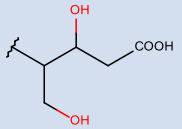
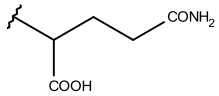
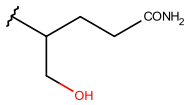
1.4 Mycosporine-like amino acids

Numerous human diseases are associated with oxidative stress (1.2.6), but multiple clinical trials of direct antioxidants have failed to demonstrate benefits to human health (1.2.8.4). Indirect antioxidants which activate Nrf2 regulated cellular defences have, however, been proven to protect against oxidative stress in mouse animal models (1.3.3) and are currently in trials for human use (112). Different natural products have the potential to activate Nrf2 regulated genes *in vivo*, and this work investigated a class of marine natural products called mycosporine-like amino acids (MAAs). MAAs are a class of UV-absorbing, low molecular mass (usually below 400 Da), natural compounds, associated with taxonomically diverse marine, freshwater and terrestrial organisms (64). MAAs are primarily UV-protective compounds, and are sometimes referred to as "microbial sunscreens" (230). These compounds have also been associated with stress-response, with evidence for antioxidant function and protection against osmotic, desiccation and thermal stresses (64,231–234).

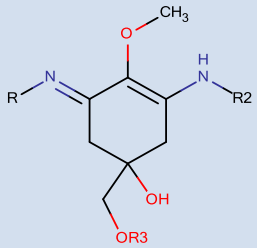
1.4.1 Structure and properties of MAAs

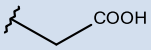
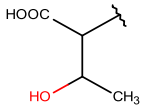
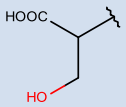
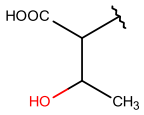
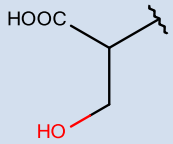
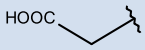
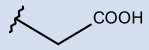
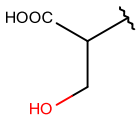
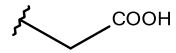
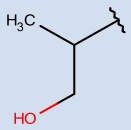
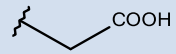
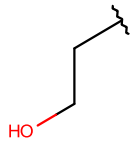
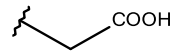
Structurally, MAAs can be classified as mono-substituted MAAs (also referred to as oxo-mycosporines) which consist of a mono-substituted amino-cyclohexenone ring (Figure 1.9/A), or as cyclohexenimine MAAs which have an amino-cyclohexenimine ring linked to an amino acid, amino alcohol or amino group (Figure 1.9/B). In addition to these structures, ester (235) and glycosylated (236,237) derivatives of MAAs have also been reported. MAA structures are reviewed in more detail in (231,232,238).

Mycosporine-like amino acids are water-soluble compounds that are transparent to visible light and absorb light in UV-A and UV-B parts of the spectrum, with absorption maxima between 260 and 370 nm. MAAs have high molar extinction coefficients (ε_{molar} = 28 100 - 50 000 depending on MAA, pH and solvent) and are photostable, harmlessly dissipating absorbed UV radiation as heat without being degraded (239,240). Due to these properties, MAAs were labelled as "The strongest UV-A absorbing compounds in nature" (241); in comparison, commercial synthetic sunscreen Avobenzone® (butyl methoxydibenzoylmethane) has a molar extinction coefficient of 40 000, while Ecamsule® (terephthalylidenedicamphor sulfonic acid) has ε_{molar} = 45 000.

A) Base structure	MAA	-R	λmax/nm	Mass/Da	MAA	-R	λmax (nm)	Mass (Da)
	Mycosporine-Gly (C10H17NO7S)		310	295.0726	Mycosporine-Ser (C11H17NO7)		310	275.1005
	Mycosporine-Tau (C9H13NO6)		309	245.0899	Mycosporine-Ser(OH) (C11H19NO6)		310	261.1212
	Mycosporine-Ala (C11H17NO6)		310	259.1056	Mycosporine-Glu (C13H19NO8)		311	317.1110
	Mycosporine-β-Ala (C11H17NO6)		310	259.1056	Mycosporine-Glu(OH) (C13H21NO7)		310	303.1318
	Mycosporine-GABA (C12H19NO6)		310	273.1212	Mycosporine-Glu(OH)$_2$ (C13H21NO8)		310	319.1267
	Mycosporine-Gln (C13H20N2O7)		310	316.1271	Mycosporine-Gln(OH) (C13H22N2O6)		310	302.1478

B) Base structure of imino-mycosporines



MAA	R	R2	R3	λmax/nm	Mass/Da
Palythine (C10H16N2O5)	-H		-H	320	244.1059
Palythine-threonine (C12H20N2O6)		-H	-H	320	288.1321
Palythine-serine (C11H18N2O6)		-H	-H	320	274.1165
Mycosporine-methylamine-threonine (C13H22N2O6)		-CH3	-H	330	302.1478
Mycosporine-methylamine-serine (C12H20N2O6)		-CH3	-H	325	288.1321
Mycosporine-2-glycine (C12H18N2O7)			-H	332	302.1114
Shinorine (C13H20N2O8)			-H	332	332.1220
Palythinol (C13H22N2O6)			-H	320	302.1478
Asterina-330 (C12H20N2O6)			-H	330	288.1321

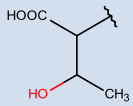
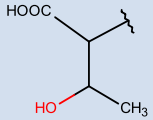
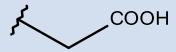
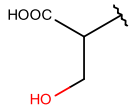
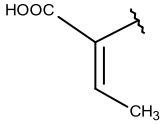
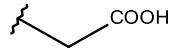
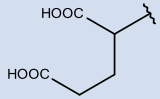
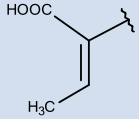
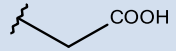
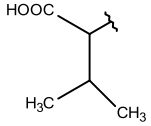
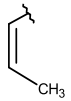
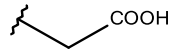
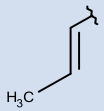
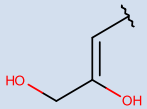
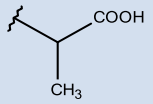
MAA	R	R2	R3	λmax/nm	Mass/Da	MAA	R	R2	R3	λmax/nm	Mass/Da
Palythine-threonine-sulfate (C12H19N2O9S)		-H	-SO3-	321	367.0817	Porphyra-334 (C14H22N2O8)			-H	334	346.1376
Palythine-serine-sulfate (C11H17N2O9S)		-H	-SO3-	321	353.0660	(*Z*) Palythenic acid (C14H20N2O7)			-H	337	328.1271
Mycosporine-Glu-Gly (C15H22N2O9)		-CH2-COOH	-H	330	374.1325	(*E*) Palythenic acid (C14H20N2O7)			-H	335	328.1271
Mycosporine-Gly-Val (C15H24N2O7)		-CH2-COOH	-H	335	344.1584	Usujirene (C13H20N2O5)			-H	357	325.1763
Palythene (C13H20N2O5)		-CH2-COOH	-H	360	325.1763	Euthalothece-362 (C14H22N2O7)			-H	362	330.1427

Figure 1.9 Chemical structures of MAAs

Figure A) displays mono-substituted (oxo) mycosporines, while Fig B) displays multi-substituted (imino) mycosporines. For each MAA, figure lists base structure of MAA, substituents (-R, -R2, -R3), maximum of spectral absorbance (λmax, in *nm*) and molecular mass in Da, rounded at four decimal points.

1.4.1.1 Role of MAAs in protection against UVR

The role of MAAs as UV protectants was established based on MAA absorbance profiles and high molar extinction coefficients (239). Studies with corals, where animals were kept at different depths underwater, found an inverse relationship between UV intensity/depth and MAA content in animal tissues (242,243). High concentration of MAAs found in UV-exposed tissues such as skin and ocular lenses of fish and other marine organisms (244) supports the UV-protective function of MAAs. In addition, research with human skin fibroblasts confirmed that the MAA porphyra-334 protects against UV-A induced photo-ageing (245) and the MAAs shinorine, porphyra-334 and mycosporine-glycine protect fibroblasts against lethal UV dose (246). Finally, a small scale human volunteer studies confirmed protection against photo-ageing using a mixture of porphyra-334 and shinorine (247). A similar MAA mixture was also observed to reduce UV-induced sunburn in an mouse animal model, as measured by clinical signs such as erythema, and biochemical parameters, such as expression of heat shock protein 70 and antioxidant enzymes (248). Altogether, multiple lines of evidence firmly demonstrated that MAAs have a UV-protective function *in vivo*.

1.4.1.2 Antioxidant functions of MAAs

A considerable part of the UV induced damage (especially for UV-A, see 1.2.2.1) is caused by the photo-induction of RS in the cells of UV-exposed tissues (40). Thus, MAAs were hypothesized to provide protection against oxidative damage complementary to blocking of UVR (232). Numerous experiments were performed to elucidate potential antioxidant properties of MAAs, and these are covered in detail in a recent review by Wada et al. (232). The topic is briefly covered here, with focus on the MAAs investigated in this work.

MAA precursors: Gadusol and 4-deoxygadusol (4-DG) were extensively tested for antioxidant activity and demonstrated to be strong antioxidants, with activity comparable to endogenous cellular antioxidants such as ascorbic acid and vitamin E. For example, gadusol was found to be a “stronger” antioxidant than Trolox (a water-soluble analogue of vitamin E), when measured by oxygen radical absorbance capacity (ORAC) assay (249); its activity is comparable to Trolox when measured for capacity to scavenge water-soluble stable radical ABTS (249). The MAA precursor 4-DG inhibited lipid peroxidation of phosphatidylcholine induced by AAPH (free radical generating compound) *in vitro* and has higher antioxidant activity than the MAAs shinorine and mycosporine-glycine (M-Gly) (64).

Mycosporine-glycine: M-Gly was found to inhibit lipid peroxidation with moderate activity (approximately 20-25% effect of ascorbic acid) (250), which was confirmed by a β-carotene bleaching method (251). Assays for free radical scavenging using the artificial stable radical DPPH and ABTS at pH 7.5 showed scavenging activity comparable to ascorbic acid (251,252).

Shinorine: A lipid peroxidation study of the MAA shinorine, based on a phosphatidylcholine peroxidation assay of shinorine-containing extracts from the ascidian *Lissoclinum patella*, found no evidence for significant activity (244,250). Later research based on a β-carotene bleaching assays, demonstrated that shinorine is effective at preventing lipid peroxidation, with activity higher then M-Gly (251). Free radical quenching assays found no activity towards DPPH radicals, and low activity against ABTS radicals, as compared to ascorbic acid and M-Gly (251,252).

Palythine: Antioxidant activity of palythine has not been extensively investigated. Early studies by Dunlap et al. (1995) found no evidence for inhibition of lipid peroxidation by palythine, as measured using the phosphatidylcholine peroxidation assay (250), while later research by Rastogi et al. (2014) detected significant antioxidant activity from a mixture of MAAs including palythine. According to this research, the antioxidant scavenging activity of this mixture was approximately 15 – 30% that of ascorbic acid as measured by DPPH, ferric-reducing antioxidant power (FRAP) and superoxide radical scavenging activity (SRSA) assays (253,254). Since these studies were not performed with purified palythine, however, further research is required to elucidate the antioxidant capacity of this MAA.

Porphyra-334: Investigation of an MAA mixture from algal extracts concluded that porphyra-334 in the mixture had low antioxidant activity, as measured by ABTS and DPPH assays (251). The same study found moderate antioxidant activity, comparable to shinorine but lower then M-Gly, when measured by β-carotene bleaching and pyrogallol superoxide generating assays (251). Multiple studies based on inhibition of lipid peroxidation have indicated that porphyra-334 protects lipids against AAPH induced peroxidation, but is less active then gadusol, 4-DG or α-tocopherol (250,252,255,256).

1.5 Bioinformatics tools for biological discovery

Over the last three decades, computational tools have played a considerable role in discovery and design of pharmaceutically active compounds, and computational approaches were recently utilized to predict small-peptide inhibitors of Keap1-Nrf2 binding (158). While extensive review of computer-aided drug design is beyond the scope of this thesis, here follows a brief introduction into methods relevant to work presented herein, with focus on approaches for reconstruction of molecular evolution (computational phylogeny) and virtual screening methods used for discovery of novel drug-like compounds.

1.5.1 Computational phylogenetics

Computational phylogenetics is a branch of bioinformatics dealing with the analysis of molecular evolution. Unlike classical phylogenetics which is based upon morphology and the study of fossil records, computational phylogeny uses nucleotide or protein sequences to infer evolutionary relationships based upon mathematical models of evolution. Results of phylogenetic analyses are commonly represented as phylogenetic (or evolutionary) trees, where edges (also called branches) represent taxonomical units, i.e. species or sequences, and nodes that connect these represent the hypothetical last common ancestors of connected taxa (257). Example of an annotated phylogenetic tree is shown in Figure 1.10.

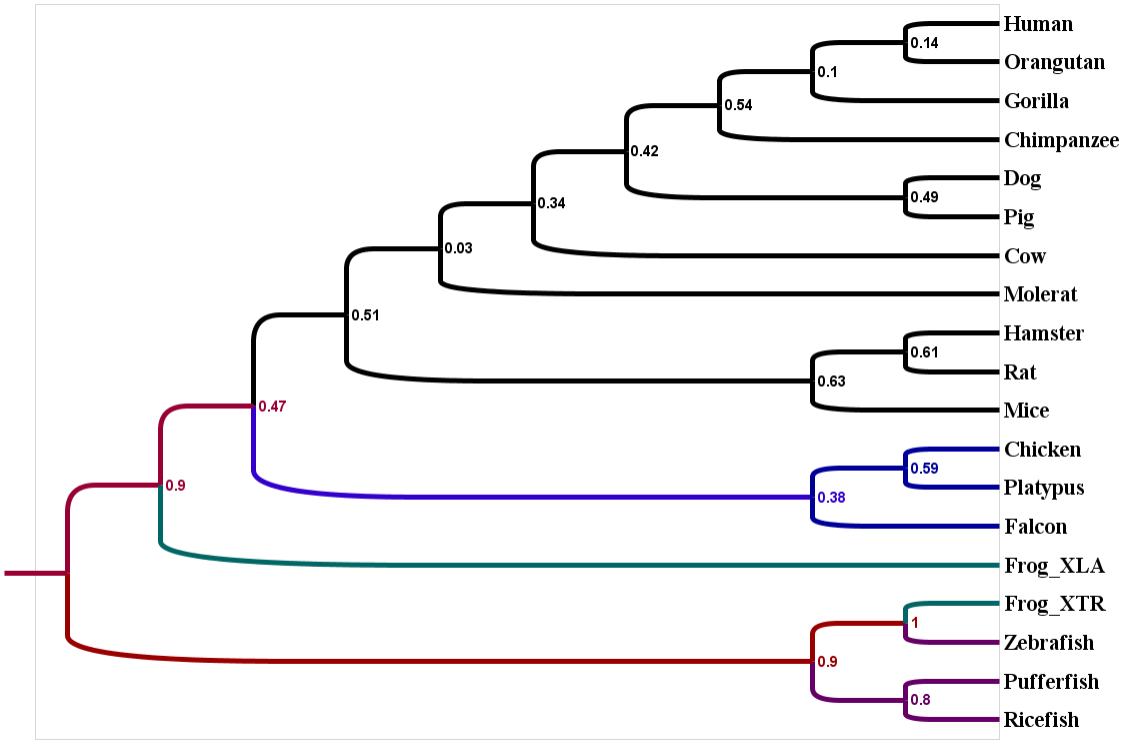


Figure 1.10 Phylogenetic tree of vertebrate Nrf2 neh2 domain

Tree shows the phylogenetic reconstruction of vertebrate Nrf2 neh2 domains. The multiple alignment of vertebrate Nrf2 neh2 domains (Figure 1.11) was generated using the T-Coffee server, and phylogenetic reconstruction was conducted using the MEGA 6.0 software. The tree was reconstructed using the maximum likelihood method, with LG substitution matrix and complete deletion of gaps. The substitution matrix was selected using the MEGA 6.0 model selection tool. Bootstrap numbers (expressed as percentage of 1000 bootstraps) are listed for each taxonomical split. The tree is colour coded, with sequences from birds coloured blue, amphibians green, fishes purple, mammals black, and polyphyletic nodes coloured red.

The phylogenetic reconstruction generates an evolutionary tree from a measure of homology, i.e. evolutionary relatedness, and mathematical model of an evolutionary process. A measure of homology is a starting point for phylogenetic reconstruction and is supplied via multiple sequence alignment (MSA) of nucleic acid or protein sequences (258).

1.5.1.1 Sequence alignment and multiple sequence alignment

A two sequence, or pair-wise, sequence alignment is a way of arranging two biological, i.e. protein, DNA or RNA sequences, defined as a list of letters (string), to identify similarities. The alignment matches identical and evolutionarily-related amino acids or nucleic acids while placing gaps in positions corresponding to likely deletion or insertion events. The multiple sequence alignment (MSA) is an extension of pair-wise sequence alignment to more than two sequences. It is typically represented as a matrix where each row is one sequence and columns

represent aligned amino-acids or nucleic acid nitrogenous bases (see Figure 1.11 for example) (259). Unlike the pair-wise alignment which was solved in the early days of bioinformatics (260,261), the optimal MSA algorithm is still an unsolved problem of bioinformatics. That is, none of currently existing MSA algorithms can generate an "optimal" alignment of many sequences, not in the least because the biologically-relevant, optimal, alignment is still to be defined (262,263).

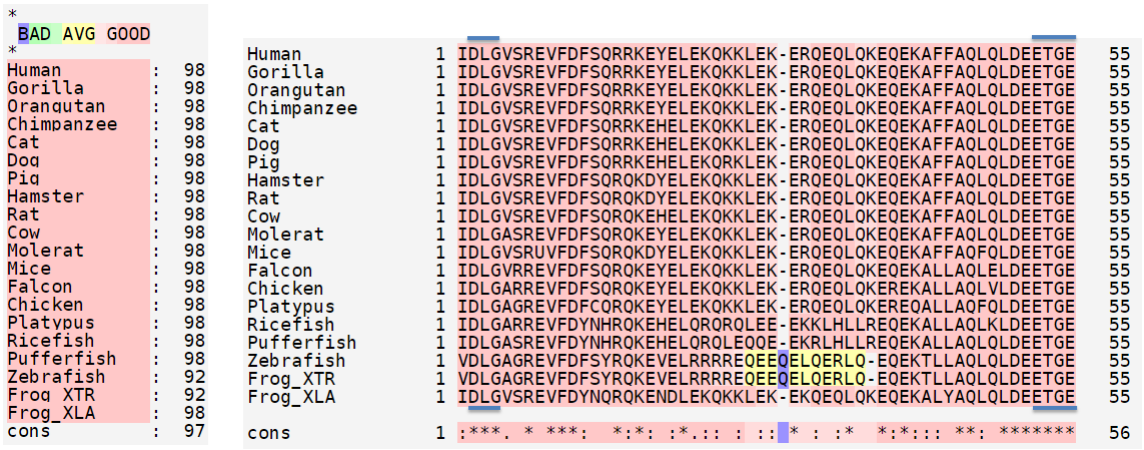


Figure 1.11 Multiple sequence alignment of Nrf2 Neh2 domains

Figure shows the example of multiple sequence alignment of vertebrate Nrf2 Neh2 domain sequences, calculated using the T-Coffee server. Alignment columns are coloured according to conservation from "bad" (blue) to "good" (pink); fully conserved columns are marked by star (*), while partially conserved columns are marked with dot (.) or colon (:). Keap1 binding motifs DLG and ETGE are marked with blue lines.

Phylogenetic study is based on the assumption that amino acids in each column of the multiple sequence alignment (MSA) are homologous (have evolved from the same position of a common ancestral sequence). Thus, MSA can be used to infer the evolution of a protein as well as structure and function (257). Active sites and other critical motifs in the protein tend to be evolutionarily conserved, as seen for example for DLG and ETGE motifs from Nrf2 in Figure 1.11. This assumption, however, only holds if the MSA represents a meaningful biological relationship; that is, phylogenetic reconstruction is only as good as the MSA used for the analysis (257). While there is no universal measure of the quality of multiple alignments, MSA quality control tools such as ProbCons (264) have been developed to assess various accuracy metrics of the MSA. In addition, consensus alignment tools such as M-Coffee allow for combinations of different algorithms to minimize the potential bias of an MSA algorithm (265), and visual inspection using tools such as JalView (266) and MEGA (267) can identify conservation and consistency of the MSA.

1.5.1.2 Methods and models for phylogenetic reconstruction

Multiple methods can be used to build a phylogenetic tree from a multiple sequence alignment; these are usually divided into *distance matrix methods* (e.g. UPGMA and neighbour-joining) and *discrete data methods,* also called tree searching methods, such as maximum likelihood, parsimony and Bayesian methods.

Distance matrix methods build a phylogenetic tree using distances between aligned sequences. These methods are based on the premise that quantification of sequence differences can be expressed as a number that provides a good model of biological/evolutionary distance. These methods calculate distance, using a method-dependent metric, for each sequence pair and assemble these into a *distance matrix.* The matrix is used to generate a tree in a step-wise fashion, by grouping the two closest sequences to form a tree node, recalculating the distance matrix, and then repeating until the tree is assembled. Different algorithms calculate the distance between sequences using different metrics, while sharing the iterative approach and the idea of quantifying evolution by a single metric. These methods are computationally fast and relatively simple to implement, but assume that evolutionary rate is identical for all amino acids in the protein (257,268).

The *discrete data methods* (e.g. maximum parsimony and maximum likelihood) examine each column of the multiple sequence alignment and search the *tree space*, i.e. all possible trees that could be composed from the input dataset, for a tree that best satisfies the criteria of the model of evolution. For example, maximum parsimony searches for the tree with least mutations that can explain the data, while maximum likelihood (ML) searches for a tree which optimizes the likelihood function. These methods can model unequal evolution rates across the protein to account for regulatory sites or critical amino-acid residues, and arguably provide more biologically relevant trees when compared to distance-based methods. These methods are, however, computationally expensive and are dependent on the choice of the model of evolution. For example, a model which assumes equal rates of evolution for all columns in the alignment can bias the results of phylogenetic reconstruction of enzyme-encoding DNA sequences, where the certain parts of sequence encode the amino acid resides that compose enzyme active sites, and evolve slower than the rest of the gene (269). As the size of tree space scales exponentially with the number of input sequences and thus, tree search methods use heuristic approaches to calculate a ‘good enough’ solution rather than the best tree (257,268). Finally, a relatively recent approach to phylogeny includes Bayesian methods; these are not unlike maximum likelihood approaches, but are based on Bayesian, as opposed to classical, statistics (268).

Models of evolution

Aside from the reconstruction method which specifies a way of treating the input MSA, phylogenetic reconstruction requires the *model of evolution* (also called *substitution model*), the mathematical description on *how* to apply a method. For example, the simplest way to describe evolution mathematically is to assume that all DNA mutations (A→T, A→G, A→C, C→G, T→G and reverse mutations) are equally likely; this model was introduced in 1969 by Jukes and Cantor and is known as Jukes and Cantor or JC69 substitution model (270). Since then, a large number of other models have been published. Detailed descriptions of various substitution models are given in (269), but briefly:

Time-reversible models such as JC69 describe the evolution by assigning a rate to nucleotide mutations and assume mutations are reversible (e.g. A→T mutation rate is the same as T→A); these can be generalized into a General Time Reversible model (GTR), which allows for a different mutation rate for each mutation (271).

Protein sequence phylogeny models are designed to explain the evolution of proteins rather than DNA sequences. Unlike DNA evolution models, these are based on observation instead of attempting to model the process of evolution, and use empirically derived substitution matrices not unlike the ones used for protein MSA. Examples include the Dayhoff model (272) and Jones-Taylor-Thornton (JTT) model (273).

Different *substitution models* have vastly different numbers of parameters and interestingly, the most complex model of evolution does not necessary produce the most biologically relevant phylogenetic tree (269). Hence, there is a need to select the model that is complex enough to explain the data, while not using too many parameters. Several statistical measures have been invented to aid in this task; for example, a common method for model selection in phylogeny is to compare Akaike's information criterion (AIC) between the models and then to select the model with the "best" (lowest) AIC value. The AIC is a measure of how well the model fits the data, taking into account the number of parameters in the model (269).

Molecular clocks and dated trees

A phylogenetic tree describes a process of evolution, but usually does not provide a correct time-frame to evolutionary events. That is, it explains ancestry of each studied taxa by virtue of taxonomical splits, but does not necessarily model *when* these splits occurred. While phylogenetic trees do calculate branch lengths, these distances are simplified metrics and are typically an inadequate representation of evolutionary history (274); also see 1.5.1.3.

Thus, reconstructing the time-frame of evolution requires application of an evolutionary or molecular clock, a hypothesis that the molecular evolution occurs at an approximately uniform rate over time (275). The "clock" is then calibrated using fossil records to assign a time-frame to one (or multiple) taxonomical splits, to generate the time-frame for the tree. This procedure is complicated by unequal rates of evolution, imperfection of fossil records (if available) and complications in the models, due to the extra parameters and computational difficulties incurred by the introduction of molecular clock (274,276); see 1.5.1.3 for more details.

1.5.1.3 Challenges and difficulties in molecular phylogeny

As described in section 1.5.1.2, phylogenetic reconstruction is a complex process that involves data selection, multiple sequence alignment, choice of a *model of evolution* and choice of a *tree reconstruction method*. Producing a time-calibrated tree also requires the *molecular clock model* and *calibration* of the tree. Here follows a brief review of common issues encountered in phylogenetic reconstruction. These issues have no single, universal solution and the *Discussion chapter* of this thesis will present the author's approaches to addressing these problems encountered during this research.

Multiple sequence alignment

Phylogenetic reconstruction is highly dependent upon the multiple alignment used to build it, and it has been shown that the choice of alignment algorithm is as important as the evolutionary model or the reconstruction method (277). A large number of MSA algorithms have been published, and whilst many share similarity in approach and can be loosely grouped into *progressive* or *iterative* aligners, different algorithms do not produce identical alignments. By illustration, a recent review lists more than 25 "commonly used" algorithms (278). Assessment of MSA algorithms is complicated as results tend to vary based on the benchmark used. For example, an algorithm optimized for short sequences might not perform well on very long sequences or sequences with large gaps. Similarly, algorithms that produce high quality alignments, such as Psi-Coffee, might lack the computational optimizations necessary to align many sequences in reasonable amount of time. Therefore, reviews of MSA tools, such as a recent benchmark by Pais et al. (2014) (279), use multiple benchmarks and rank the tools depending on the test used.

Evolutionary model and tree-reconstruction method

In addition to the choice of MSA algorithm, the choice of model and reconstruction method also form critical part of phylogenetic reconstruction (277). Selection of the model, however, is arguably a less difficult problem due to the existence of statistical frameworks for model selection. For example, AIC and Bayesian statistics equivalent (BIC) provide a measure of how well the model fits the data (269).

Establishing an evolutionary time-frame

Establishing the correct time-frame for a phylogenetic tree requires a molecular clock hypothesis, that is, an assumption that evolution progresses in predictable, clock-like, manner. This is, however, complicated by the fact that there is no universal molecular clock, because the different DNA sequences evolve at different rates (276). For example, most microorganisms have very short generation time and thus evolve faster than large animals with long generation time (280). Other factors, such as difference between sexual and asexual reproduction, impact of mutation to organism's reproductive success, genetic drift, and metabolic rate of the studied organism have been associated with the rate of evolution (281). A non-constant molecular clock is not a major problem when phylogenetic analysis is performed for closely related species, similar genes, or short time periods, but it complicates the dating of phylogenetic trees which span long time intervals, include very diverse species or have to account for potential changes in a gene functions (276,282).

Construction of the correct time frame for a phylogenetic tree also requires one (or more) calibration dates, usually obtained from ages of known fossils. The, fossil record is however biased towards organisms susceptible to fossilization (large organisms with hard skeletons resistant to decay), and the fossils of organisms that evolved before the Cambrian explosion and evolution of major animal phyla are scarce (283,284).

1.5.2 Virtual screening for drug discovery

Virtual screening (VS) uses algorithms and computer programs to evaluate how likely a small molecule is to bind to a drug target, typically an enzyme or receptor protein. Somewhat similar to *in vitro* high throughput screening, VS is an *in-silico* approach in early stage drug design and discovery, and can be used to evaluate a large number of chemical structures to identify candidate compounds for follow-up *in vitro* studies. VS approaches can be divided into two broad categories, *ligand-based screening* and *structure-based screening* (285,286). Figure 1.12 illustrates common virtual screening approaches and illustrates the main differences between *ligand-* and *structure-* based screening. It should be noted, however, that while often treated as alternate approaches to *in-silico* prediction of novel leads, *structure* and *ligand*-based approaches are not mutually exclusive and have been utilized complementarily (287,288).

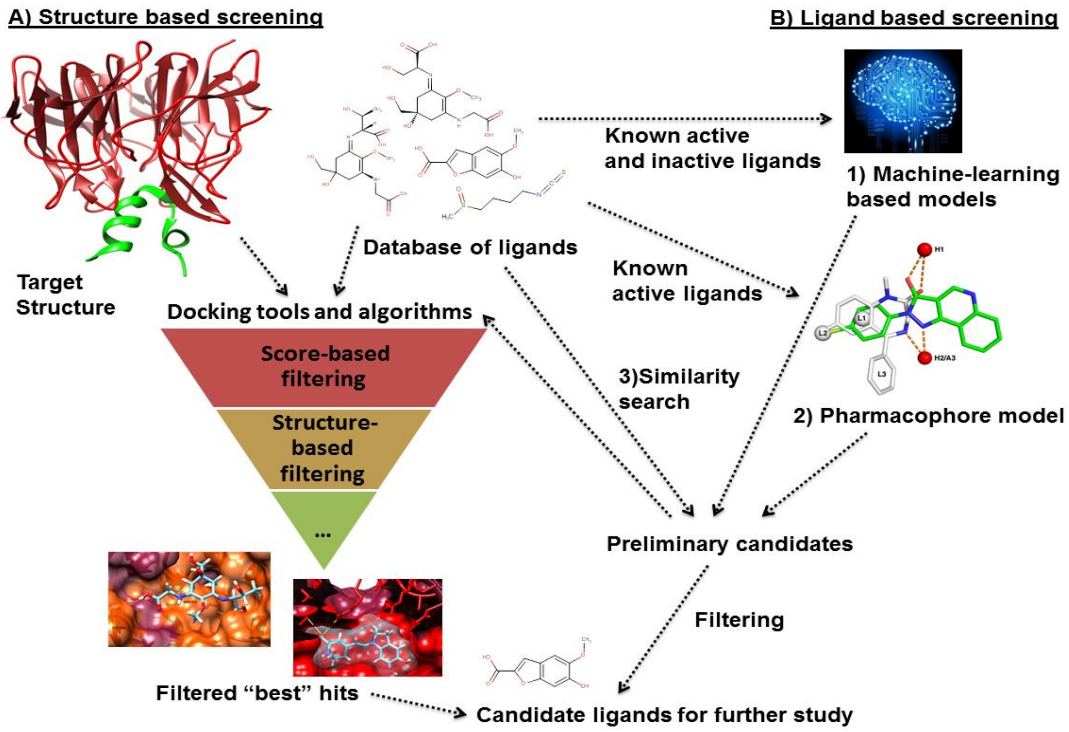


Figure 1.12 Virtual screening

Schematic illustrates common virtual screening approaches. Structure based screen (A) is based on known 3D structure of the target and involves computational docking of large number of ligands onto the target; ligands are typically filtered based on docking scores and predicted chemical properties, e.g. hydrogen bonds, and examined manually to select "best hits". Ligand based screening (B) approaches are based on: 1) examination of known active ligands to produce a pharmacophore model, 2) machine learning models built from known active and inactive ligands, and 3) searching chemical databases to identify ligands similar to known active compounds. These models are used to predict candidate molecules which are then further filtered based on chemical properties or druglikeliness, e.g. by Lipinski's rule of five. Notably, these approaches can be combined to generate list of candidates for further study. Schematic based on (285,289)

1.5.2.1 Ligand-based screening

A ligand-based approach uses the information extracted from the known target-binding ligands to predict novel chemical leads, with the aim to predict compounds with better affinity for the target or other desirable, "drug-like", properties (290) such as lower molecular mass, or lower polarity. These variants of virtual screen can be conducted in a number of ways, depending on the available data; for example, knowledge about the binding-affinity of multiple ligands allows the construction and training of machine-learning based models of ligand-target interaction (291); alternatively, ligand structures and activities can be used to construct a pharmacophore or a quantitative structure-activity relationship (QSAR) model (292). If less information about ligand-target interactions is available, the knowledge of a few active ligands is enough to produce a less sophisticated pharmacophore, i.e. model of molecular recognition by ligands (293). Finally, the knowledge of at least one active ligand allows the search for similar chemical structures. These approaches are used individually or combined to predict novel compounds which are likely to bind to the target. Notably, none of these approaches require the 3D structure, e.g. x-ray crystallographic model, of the target (285,286).

1.5.2.2 Structure-based virtual screening

In contrast to the ligand-based screening, structure-based approaches start with the 3D structure of the target, e.g. enzyme or receptor protein, and use computational docking simulations to determine the binding affinity of test compounds, calculated using a scoring function. Such screening is often performed with large libraries of compounds, and studies with hundreds of thousands of candidate compounds are not unusual (see (294–296) for few recent examples). The results of the screen are filtered in step-wise fashion, e.g. by docking score, followed by clustering to identify high-scoring scaffolds, drug-likeliness analysis, and finally manual examination of docking models. Numerous compound libraries and tools are available for docking analyses, and examples of free-for-academic-use tools and resources are listed in Table 1.4; more comprehensive lists are available in recent reviews, e.g. (289,297,298).

Table 1.4 Commonly used ligand databases and docking tools

Table lists examples of commonly used ligand libraries and docking tools for structure-based virtual screening. Listed tools are free for academic use, well documented and widely used in the academia.

Ligand library	Notes	Resource
ZINC database	Over 100 million compounds, docking ready	http://zinc15.docking.org/
PubCHEM	Over 200 million compounds, also contains data collected from over a million bioassay experiments.	https://pubchem.ncbi.nlm.nih.gov/
DrugBank	∽8200 drugs, ∽21000 drug-targets. Contains extensive annotation of drugs and targets.	http://www.drugbank.ca/

Docking tool	Notes	Resource
AutoDock 4.2	Supports flexible and rigid-body docking, extensive documentation and GUI integration (ADT, Raccoon) are available.	http://autodock.scripps.edu/
AutoDock Vina	Supports flexible and rigid-body docking, documentations, tutorials and GUI integration (Raccoon) are available.	http://vina.scripps.edu/
UCSF DOCK	Supports flexible and rigid-body docking, multiple scoring functions; USCF Chimera plugins allow limited graphical interface.	http://dock.compbio.ucsf.edu

1.5.2.3 Considerations and limitations of virtual screening

Structure-based VS is considered more effective at finding new chemical scaffolds than the ligand-based approach (299), but it requires a high-quality 3D structure of the target. These are usually a result of x-ray crystallography studies, which tend to be resource-intense and exceedingly specialized, and thus most virtual screens rely on structures deposited in the PDB database (300). The lack of structure can, in principle, be rectified by homology modelling, i.e. prediction of protein structure based on 3D structures of similar proteins (templates), but protein structure prediction is of little utility when templates for homology modelling are not available (301). *De-novo* protein modelling, i.e. prediction of 3D structure from protein amino-acid sequence, is a potential solution when no templates are available, but it is a challenging endeavour unlikely to produce a high-resolution 3D structure, especially when applied to large proteins (302,303). In addition to the need for a high resolution model of the target, structure-based screens are computationally expensive, and often employ specialized, high-performance computing, hardware and can produce different results depending on the tools and algorithms used (304).

Finally, there is a matter of the accuracy and biological relevance of the VS approach. Structure-based screens use simple and crude approximations of biological systems, and while

different tools use varied approaches, in general these screens treat a target macromolecule as a rigid body and ligands as rigid or semi-flexible (304). Determination of docking pose is based on scoring functions, which are often empirically or knowledge derived. In addition, these methods treat the system as static, and the movement of target and ligands during the docking is not included in the model, and do not model chemical reactions. These approaches are necessary as full quantum-chemistry treatment of large systems such as proteins is not yet precise/feasible and even simpler, molecular-dynamics based, approaches require very high computation time and are therefore impractical for the testing of large number of ligands (304,305).

Benchmarking the performance of VS and tools for screening is a difficult prospect, as definition of success is not straightforward. For example, a screen that successfully finds new interesting chemical scaffolds with low hit rate is arguably "more successful" than a screen which detects large numbers of ligands with known activities. In addition, benchmarks of VS are retrospective, i.e. based on a method's performance in retrieving known active ligands, and thus cannot be considered a good measure of tool's performance in finding new, chemically different, ligands (306–308).

1.6 Multidimensional protein identification technology

Multidimensional protein identification technology (MudPIT) is a technology for identification of the protein content in a biological sample. Also called shotgun proteomics, MudPIT is based on a combination of protein separation, digestion of the proteins by proteases, mass spectrometry, and bioinformatics for data analysis. Proteins are roughly separated into a number of fractions, usually by electrophoresis, and digested into small peptides, e.g. by trypsin. The resulting peptides are further separated using HPLC and introduced into the mass spectrometer using "soft-ionization" techniques such as ESI or MALDI to obtain mass spectra (MS) of peptide ions and tandem mass spectra of fragmented ions (MS/MS). The analysis produces a large number of mass spectra (typically in tens or hundreds of thousands) which are computationally processed and identified, commonly by database matching algorithms. Algorithms compare mass spectra to theoretical spectra generated by *in-silico* fragmentation of a database of proteins from the subject organism or closely related organisms (309–311). An example of MudPIT workflow is presented in Figure 1.13.

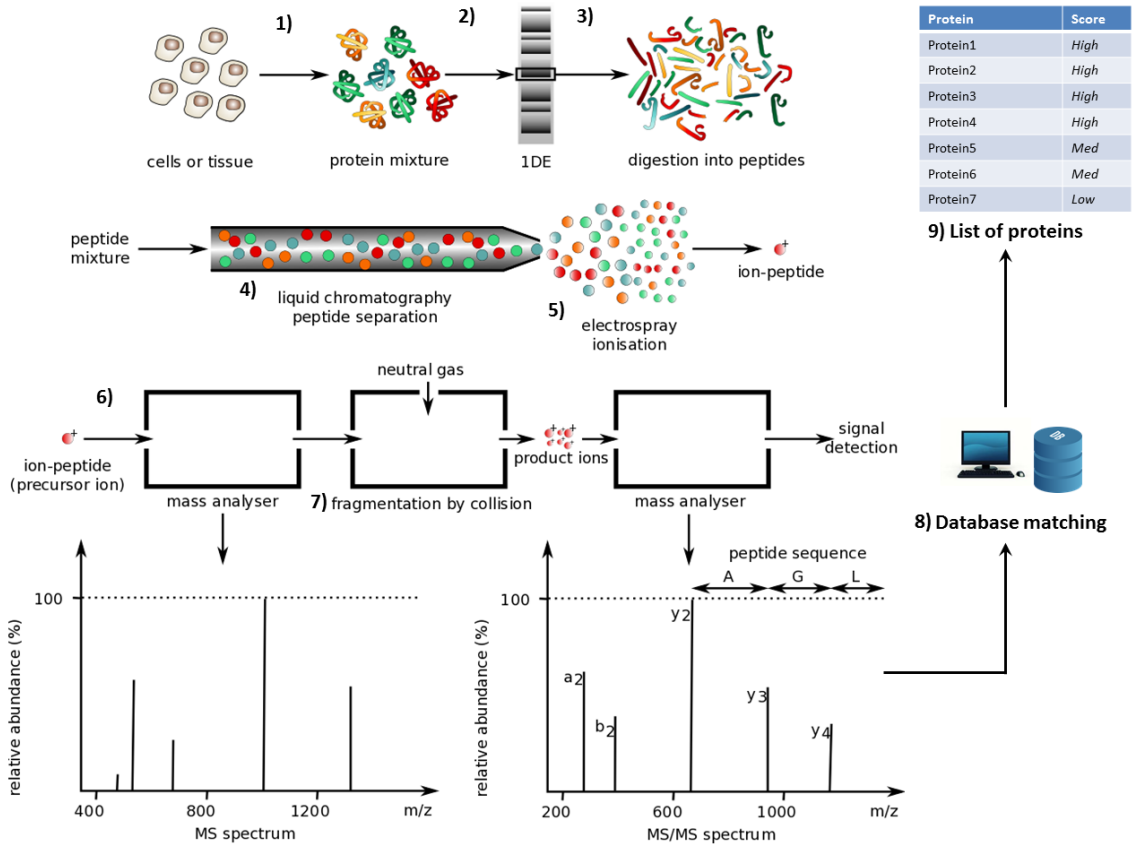


Figure 1.13 MudPIT workflow schematic

Schematic illustrates an example of MudPIT proteomics experiment. Proteins are extracted from the sample (1), separated using gel electrophoresis (2) and digested into peptides, e.g. by using trypsin (3). The peptides extracted from the gel are separated via HPLC (4) and introduced into mass spectrometer by soft ionization technique such as ESI (5). Mass spectrometer measures mass spectrum of each peptide ion and tandem mass spectrum produced by peptide ion fragmentation (7). Finally, these spectra are matched to database of theoretical spectra (8) to produce list of detected proteins (9). Figure adapted from (312).

1.6.1 Quantitative and qualitative proteomics

While originally introduced as a qualitative method that identifies proteins but provides no information about expression levels (313), the shotgun proteomics workflow has since been enhanced to enable quantification of protein expression. The protein expression can be quantified with chemical labels, or calculated from mass spectra by utilizing label-free methods (314).

1.6.1.1 Label-free quantification

Label-free quantification methods compare the mass spectra (MS) generated from identical proteins in different samples to calculate the relative protein expression levels. Two approaches have been used in this type of quantification; the first approach calculates expression from peaks in the mass spectra, using peak areas or peak heights, and the other approach is based on counting the number of spectra identified for the same protein in different samples (314). While attractive as a concept, these methods are complicated by several issues that limit applicability in experimental settings. For example, peptide ion fragmentation is not consistent between different peptides or proteins, and different peptides present in the same concentration do not generate ion peaks of the same intensity (315). A recent study by Schweiki et al. (2017) benchmarked different label-free quantification methods on the proteome of the plant *Arabidopsis thaliana* (316). The authors identified that label-free methods are inaccurate when used to quantify short proteins and low concentration proteins which generate few peptides and mass spectra. The study also concluded that label-free methods lack the sensitivity required to correctly quantify proteins with expression fold-changes below 2.0 (316).

Nonetheless, label-free methods have been employed successfully, and are a promising approach due to simplicity (no extra reagents are required for the experiment) and the ability to compare a large (theoretically unlimited) number of samples (314). For example, Patel et al. (2011) used label-free approach to compare protein expression of 55 tissue samples obtained from chronic hepatitis C patients (317). A review by Megger et al. (2013) identified 15 different label-free proteomics studies published between the years 2011 and 2013, and illustrated that the label-free approach is well-utilized in clinical studies (318).

1.6.1.2 Quantification by metabolic and chemical labelling

Quantification by metabolic labelling, such as stable isotope labelling with amino acids in cell culture (SILAC), is based on the assumption that biochemical properties of proteins do not change if amino acids in the protein are replaced by "heavy" amino-acids such as Lys or Arg containing stable isotopes ^{13}C or ^{15}N. The cells cultivated in the medium with "heavy" amino acids incorporate these into proteins, and replace the majority of the proteome with "heavy" proteins over five generations. Labelled cells are then combined with unlabelled cells and processed according to the standard MudPIT workflow. The peptides from labelled cells have higher mass than corresponding "light" peptides that contain normal amino acids, and the shift in the mass can be calculated from the peptide sequence (for example, shift is 6 Da in the case of $^{13}C_6$-Arg). The mixture of "light" and "heavy" peptides generates the mass spectra with two peaks – one from ionization of the "heavy" peptide and the other from "light" peptide ion.

Because the peptides are chemically identical, except for difference in mass, the ratio of ion peak intensities equals the ratio of the protein concentration in "heavy" sample versus the "light" sample. This enables the relative quantification of protein expression between the samples (314,319). An example of SILAC workflow is illustrated in Figure 1.14/A.

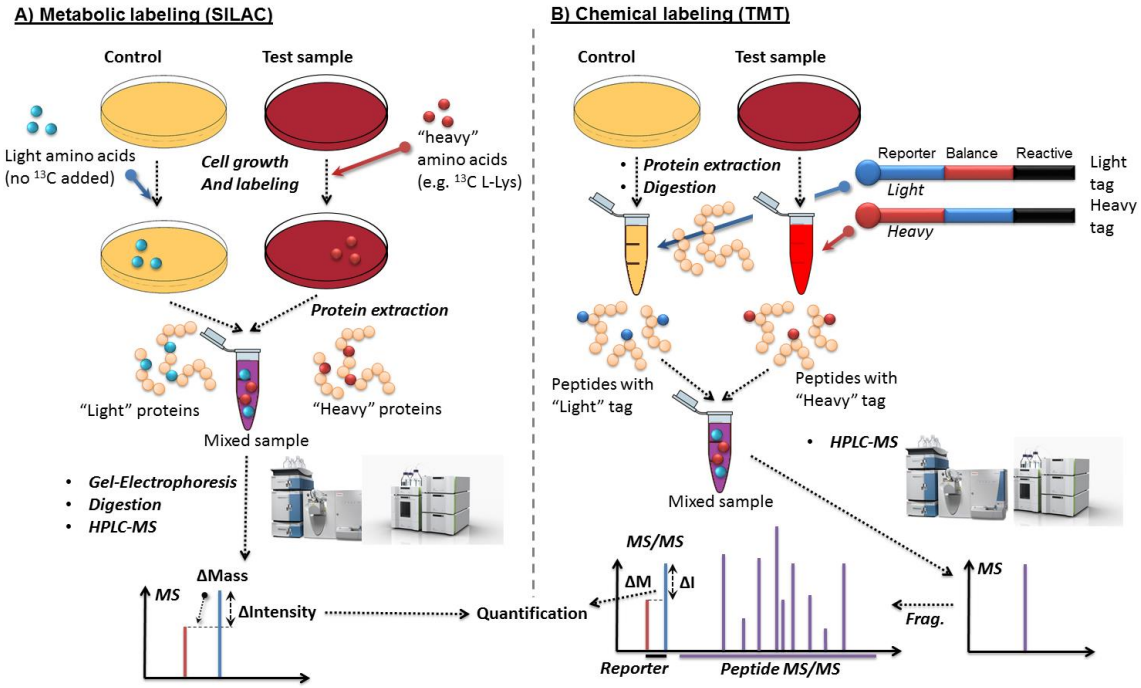


Figure 1.14 Comparison of metabolic and chemical labelling

Schematic displays workflows for metabolic labelling (A) and chemical-tag based labelling using TMT tags (B). Metabolic labelling, e.g. SILAC, is based on adding a "heavy", i.e. ^{13}C substituted, amino acids into growth media; these are incorporated into the cellular proteins metabolically and quantification is based on spectral shift in MS spectrum of resulting peptides. Chemical labelling by TMT is based on labelling of peptides after protein digestion, e.g. by trypsin; TMT "heavy" and "light" tags have identical masses and are not observed in MS spectrum of a peptide, but have different reporter region masses; tags fragment during MS ion fragmentation and reporter ions are used to quantify peptides from MS/MS spectrum.

Chemical labelling for protein quantification is based on attachment of chemical "labels" or "tags" to proteins or peptides isolated from test and control samples. A large number of different chemical tags have been developed since the introduction of high-throughput proteomics, and these are described in detail in recent reviews (314,320,321). All chemical tagging methods use similar approaches which are described here, using tandem mass tags (TMT) labelling as an example, (322). Quantification, illustrated in Figure 1.14/B, is based on attachment of a TMT label to the peptide N-terminus. The label is composed of three parts: a reactive group that attaches to the peptide, a reporter ion that is registered in the mass spectrum generated by peptide-ion fragmentation (MS/MS), and a balance (or mass normalization

group). A TMT kit includes 2 to 10 (depending on the kit type) different chemical labels with identical total masses but different reporter ion masses (counterbalanced by mass normalization groups to achieve the identical total mass). These kits enable the relative quantification of proteins from up to 10 samples (323). The TMT label reporter ion group fragments during the secondary fragmentation of the peptide ion in the mass spectrometer, after peptide ion spectrum (MS) is recorded, but before the identification of MS/MS spectrum. Thus, the differentially labelled peptides have identical MS spectra, but generate reporter ions with different masses in MS/MS spectra. As intensities of reporter ions of peptides from different samples correspond to abundancies of these peptides, relative quantification is achieved by comparison of reporter ion intensities, while the peptides are identified, e.g. by database matching, based on the MS/MS spectrum generated from the labelled peptide (322).

1.7 In summary

The section 1.2 of this introduction briefly reviews the concept of oxidative stress, its causes and consequences, and presents some of the controversies of redox biology. The oxidative stress, introduced in section 1.2.1 and discussed in sections 1.2.2 - 1.2.7, is caused by the increase in cellular concentration of free radicals and other reactive chemical species (RS). This causes oxidative damage to the cell, and can lead to apoptotic or necrotic cell death. The sources of RS, described in 1.2.2, include the mitochondrial respiratory chain as the major source of RS within the cell, and external sources such as UV radiation, pollutants and environmental toxins. As reviewed in section 1.2.5, the RS are routinely neutralized by cellular enzymatic and non-enzymatic antioxidant systems, and the cell has a capacity to tolerate and repair RS-caused oxidation. RS overload, however, leads to oxidation of all major cellular components including DNA, proteins and lipids. The cellular damage caused by the oxidative stress has been associated with neurodegenerative disorders, with diseases such as atherosclerosis and diabetes, and with processes of carcinogenesis and ageing (1.2.7).

Despite the association of oxidative stress with disease, RS are not purely deleterious, and certain RS, such as H_2O_2, are also involved in cellular signalling. The recently postulated "redox code" model states that the cell contains a "redox signalling network" composing a large number of redox-sensitive signalling proteins that are controlled by common "control nodes" such as GSH and thioredoxin. According to this model, oxidative stress cannot be explained only by an increase in RS levels, and is also caused by dysregulation of the "redox code". The concept of radical-free oxidative stress is described in section 1.2.4.

Section 1.3 describes the Keap1-Nrf2 pathway and its role in response to oxidative stress, health and diseases. The transcription factor Nrf2, described in detail in 1.3.2, is a major component of the animal "redox code"; it is a modular protein composed of seven functional domains (Neh1 – Neh7), each of which plays a distinct role in regulation of the nuclear concentration of Nrf2. It regulates, by binding to antioxidant response element (ARE), the transcription of numerous genes involved in detoxification and production of cellular antioxidants, described in 1.3.3. The cellular concentration of Nrf2 is primarily controlled by Keap1 protein which binds to the Neh2 domain of Nrf2 to sequester it in the cytosol for ubiquitination and degradation. The Keap1 contains a number of "RS sensing" cysteine residues and its binding to Nrf2 is inhibited during oxidative stress to allow nuclear translocation of Nrf2 that results in upregulation of cellular defences to neutralize the excess RS. In addition to Keap1 dependant regulation (1.3.2.3), activity of Nrf2 is also controlled by Keap1-independent ubiquitination and degradation by β-TrCP, by other signalling proteins such as p21 and p62, by microRNAs and epigenetic mechanisms. These mechanisms are reviewed briefly in section 1.3.2.4.

As reviewed in 1.3.3, Nrf2 has been implicated in protection against oxidative stress and disorders associated with oxidative stress. Numerous cell-based and mouse model studies have identified that Nrf2 knockout models are highly sensitive to carcinogens such as benzo[a]pyrene, drug toxicity (e.g. to acetaminophen and cisplatin) and environmental pollutants such as cigarette smoke. In addition, treatment with Nrf2 activators such as sulforaphane (SFN) demonstrated to protect mouse models and cell lines against oxidative stress induced by drugs or UV irradiation. Yet, while Nrf2 is essential for induction of cellular defences, analysis of tumours showed that Keap1-mediated degradation is dysfunctional in certain types of cancer, leading to enhanced transcription of Nrf2 regulated genes. These types of cancer are found to be highly aggressive and resistant to chemotherapeutics, and this phenomena has been dubbed "the dark side of Nrf2" (reviewed in 1.3.4). The dysfunction of Keap1-mediated degradation of Nrf2 was also found to be lethal in mouse models, and induction of Nrf2 regulated genes was implicated in skin pathologies and acceleration of late-stage atherosclerosis (in mouse models). As evident from the "dark side of Nrf2" examples in cancer, atherosclerosis and skin diseases (1.3.4), Nrf2 plays complex role in animals and its upregulation is not always beneficial for the organism.

Numerous diseases are associated with oxidative stress (1.2.6), but antioxidant therapies with direct antioxidants, such as ascorbic acid or α-tocopherol, have not been successful in preventing or curing oxidative stress-related diseases. The failure of clinical trials of

antioxidants is reviewed in 1.2.8, along with controversies in definitions of oxidative stress and antioxidants, and arguments against the causal link between oxidative stress and human diseases. As discussed in 1.2.8.4 and 1.3.3, activation of Nrf2-regulated cellular defences by indirect antioxidants such as SFN or other natural and synthetic compounds is a potential novel therapy for diseases caused by oxidative stress. Introduced in section 1.4, mycosporine-like amino acids (MAAs) are natural products involved in the protection of marine life and terrestrial microorganisms against UV radiation. In addition to being "microbial sunscreens", MAAs are implicated in protection against heat shock, desiccation and oxidative stress (1.4.1).

Sections 1.5 and 1.6 introduce methods and approaches used in this work. Computational methods for biological discovery are described in the section 1.5. These include methods for reconstruction of evolutionary relationships (computational phylogeny, 1.5.1) and virtual screening approaches for prediction of novel receptor-binding small compounds (1.5.2). The multidimensional protein identification technology (MudPIT), reviewed in 1.6, combines HPLC based protein separation, mass spectrometry and bioinformatics for identification of mass spectra to enable high-throughput identification of proteome. Combined with protein labelling techniques (1.6.1.2), MudPIT allows for the relative quantification of protein expression, and enables the study of proteome-wide response to oxidative stress or indirect antioxidants.

1.8 Research aims and objectives

The primary goal of this project was to identify a Keap1-Nrf2-ARE pathway in microorganisms and basal metazoans, and to reconstruct the evolution of Nrf2 in these organisms. A secondary goal of the project was to identify novel low molecular weight, natural product, Nrf2 activators produced by microorganisms. The project was based on the hypotheses that:

1. The vertebrate Nrf2 signalling pathway evolved from the simple progenitors such as microorganisms or basal metazoans. The evolution of Nrf2 signalling was driven by currently unknown selective pressures.
2. Secondary products of microorganisms have the potential to regulate the transcription of Nrf2 controlled genes in vertebrates
3. A microorganism based Nrf2 activation assay can provide a feasible alternative to assays using animal or human cells.

Chapter 2: Bioinformatics analyses provide insight into distant homology of the Keap1–Nrf2 pathway

The content of this chapter was published as:

Gacesa R., Dunlap W.C., Long P.F. (2015). **Bioinformatics analyses provide insight into distant homology of the Keap1–Nrf2 pathway.** Free Radic Biol Med. 88(PtB):373-80.

R.G. designed and developed the software used to identify distant homologs of Keap1 and Nrf2 proteins in animals and fungi, assembled the data, performed the phylogenetic reconstruction of Keap1 and Nrf2 sequences, performed the virtual screening simulations, interpreted the results and drafted the manuscript.

Supplementary material for this chapter is included as Appendix A. In-detail list of protein sequences used in the phylogenetic reconstruction is included in the electronic format on the *Appendix Disk* as *Appendix A* and is also available as online article supplement at http://www.sciencedirect.com/science/article/pii/S089158491500283X.

2.1 Foreword to Chapter 2

As reviewed in section 1.3, numerous studies have demonstrated the importance of Nrf2 in the regulation of cellular response to oxidative stress. The Keap1-Nrf2 pathway is well described in mouse animal models and human cell-line models, and known to exist in vertebrates and some invertebrates such as worm *Caenorhabditis elegans* (324) and fly *Drosophila melanogaster* (325). However, the presence of Nrf2 and Keap1 proteins in eukaryotic genomes has not been systematically evaluated, and it is currently unknown whether genomes of basal metazoans (such as cnidarians) or microorganisms contain homologs to vertebrate genes encoding Keap1 and Nrf2 proteins.

The work presented in this chapter is based on the hypothesis that microorganisms contain Nrf2 or analogous pathway, and that Keap1-depandant inhibition of Nrf2 is retained across taxonomically divergent phyla. As such, it was postulated that microorganisms produce endogenous activators of vertebrate Nrf2. Mycosporine-like amino acids (MAAs), described in section 1.4, are secondary metabolites of cyanobacteria and certain fungi known to serve various protective functions in taxonomically diverse organisms including vertebrates. Considering these compounds are UV protectants and associated with resistance to desiccation,

heat shock and oxidative stress, it was postulated that MAA are potential activators of Nrf2 regulated genes, and exert their multipurpose protective functions in part by activating Nrf2 regulated cellular defences. This chapter presents results of comprehensive phylogenetic analysis of Nrf2 in large number of currently sequenced eukaryotic genomes, and virtual screening based evaluation of MAAs for potential indirect antioxidant activity via competitive inhibition of Keap1-Nrf2 binding.

2.2 Abstract

An essential requirement for the evolution of early eukaryotic life was the development of effective means to protect against metabolic oxidative stress and exposure to environmental toxicants. In present-day mammals, the master transcription factor Nrf2 regulates basal level homeostasis and inducible expression of numerous detoxifying and antioxidant genes. To examine early evolution of the Keap1-Nrf2 pathway, we present bioinformatics analyses of distant homology of mammalian Keap1 and Nrf2 proteins across the Kingdoms of Life. Software written for this analysis is made freely available on-line. Furthermore, utilizing protein modelling and virtual screening methods, we demonstrate potential for Nrf2 activation by competitive inhibition of its binding to Keap1, specifically by UV-protective fungal mycosporines and marine mycosporine-like amino acids (MAAs). We contend that co-evolution of Nrf2-activating secondary metabolites by fungi and other extant microbiota may provide prospective compound leads for the design of new therapeutics to target activation of the human Keap1-Nrf2 pathway for treating degenerative diseases of ageing.

2.3 Introduction

The emergence of oxygenic photosynthesis, evolved first by proto-cyanobacteria approximately 3.4 billion years ago, gave rise to the Earth's oxygen atmosphere rendering subsequent progression to eukaryotic and metazoan life possible (326). Such an oxidative environment, however, posed a significant challenge to early life forms, requiring effective means of oxidative cytoprotection. In mammals, the Kelch-like ECH-associated protein 1 (Keap1) forms a complex with the nuclear factor erythroid 2-related factor 2 (Nrf2). The Keap1-Nrf2 complex dissociates in response to reactive oxygen species (ROS), releasing Nrf2 that binds to the nuclear antioxidant response element (ARE) to coordinate transcription of multiple antioxidant, detoxifying and cell survival genes (67,327). Belonging to the 'cap-n-collar' family of transcription factors that have a distinct basic leucine-zipper motif (328), the domain elements of Nrf2 are highly conserved across many diverse species, with orthologs having been detected in *Caenorhabditis elegans* (SKN-1) (329), *Drosophila melanogaster* (Nrf2-like) (325) and yeast (Yap1) (330). A prokaryotic homolog of Nrf2 (possibly OxyR or SoxR) (331,332) has also been suggested to protect UV-tolerant bacteria by augmenting coenzyme Q reduction via activation of cellular NAD(P)H: quinone oxidoreductase (NQOR) (333,334). We contend that early adaptive features of the Keap1-Nrf2 pathway conserved in extant microbiota may serve as a novel pharmacomimetic model for the discovery of new therapeutic activators of the human oxidative stress response, such may retard the progression of age-related degenerative disease, stimulate the innate immune response and suppress carcinogenesis (154,335–337). Accordingly, a new bioinformatics conduit to search and map distant homology has been developed and, in addition, Bayesian inference methods have been used to construct phylogenetic trees of Keap1-Nrf2 evolution across major eukaryotic taxa. A protein model and virtual screen were also established to predict likely activation of the Keap1-Nrf2 pathway utilizing a library of structurally diverse natural products (338,339).

2.4 Materials and Methods

2.4.1 Data retrieval

Custom databases of archaeal, bacterial and fungal proteins were constructed from the National Center for Biology Information (NCBI) Non-Redundant (NR) database (340) and the NCBI Taxonomy database (341). Sequences of human Keap1 and Nrf2 proteins, together with known homologs and predicted orthologs, were acquired from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (342) and are provided in Appendix A-1. A novel distant

homology search pipeline (called DHSP) was developed to increase the sensitivity and precision of distant homology searches by utilizing multiple Hidden Markov models (HMMs). The pipeline, described in Appendix A-2, and freely available for download at https://github.com/rgacesa/DHSP, performs a psi-BLAST (343) sequence alignment search against the NR database to detect close homology to generate HMM models. DHSP uses HMMER3, a HMM based sequence alignment tool (344), for high sensitivity distant homology detection. To minimize false positive hits, DHSP performs searches using multiple HMM models employing the Smith-Waterman algorithm (261) to align potential distant homologs against known sequences, thereby filtering out those sequences that fail to align. A tool for mapping distant homologs against the NCBI taxonomic database was additionally developed. This tool, called Taxonomy Landscape Mapper (TLM) displays results in a user friendly visual format as described in Appendix A-3. TLM is made freely available for use at https://github.com/rgacesa/TLM.

2.4.2 Phylogenetic reconstruction of Keap1 and Nrf2 homology

Multiple alignments of Keap1 and Nrf2 homologs detected by DHSP and TLM were constructed utilizing ClustalW2 (345). Phylogenetic reconstruction of Keap1 and Nrf2 proteins were assembled using the MrBayes 3.2 Bayesian inference analysis tool (338) with a mixed model (*aamodelpr=mixed*) for automatic estimation of the amino acid matrix during a Metropolis-Hastings Markov Chain Monte Carlo (MCMC) simulation, which was run for 2,000,000 generations using 25 chains. Post-run examination of parameters indicated convergence to the JTT amino-acid substitution model (with posterior probability of 100 %). Among-site rate variation was set to the gamma model with 4 categories. Other parameters of MrBayes run were left at default values. Probability and tree summaries were inspected manually to confirm simulation convergence, and─all simulations resulting in "good convergence" were accepted if the average standard deviation of split frequencies was < 0.01 with a convergence value (Potential Scale Reduction Factor) approaching 1.0 (346). Phylogenetic trees were inspected and edited using Archaeopteryx (347) and MEGA 6.1 (267) software tools.

2.4.3 Virtual screen for competitive inhibitors of Keap1-Nrf2 binding

The published data (348) extracted from the Protein Data Bank (PDB) (349) entitled "Crystal Structure of the Kelch-Neh2 Complex" (PDB-2FLU, http://www.rcsb.org/pdb/explore.do?structureId=2flu) was used to construct a virtual compound screening model to predict the release of Nrf2 by competitive inhibition of Keap1-Nrf2 binding. A model of the human Keap1-Nrf2 interaction was prepared for UCSF DOCK 6.0 (350) and AutoDock Vina (351) virtual screening algorithms using the listed protocols.

2.4.3.1 Model preparation for DOCK 6.0 screening

Receptor preparations were performed using the standard protocols for DOCK 6.0 as follows:

a) The selected protein was cleared of any artifacts, water, bound ions, secondary chains and ligands using the UCSF Chimera tools.
b) The Chimera *Dock Prep* tool was used to prepare the receptor for docking by adding polar hydrogens, applying partial charges, removing solvent and non-complexed ions and fixing any errors in amino-acid residues by using the residue conformations library (352).
c) A processed model was used to generate a molecular surface using the Chimera *Write DMS* tool.
d) Potential docking areas ("spheres*"*) were determined using the DOCK 6.0 *sphgen* tool. Sphere clustering was performed by assessing the proximity to the ligand where possible (using *sphere_selector*), and by a combination of clustering by *sphgen* and manual inspection by *showsphere* when a ligand-receptor complex is not available.
e) The DOCK 6.0 tools *showbox* and *grid* were used to generate boundaries of potential docking areas ("gridbox").
f) Models, spheres and gridbox were inspected using Chimera, and corrections to parameters (sphere clusters and sizes, gridbox size) were made to insure the use of high quality docking parameters.

2.4.3.2 Model preparation for AutoDock Vina screening

Receptor model preparations were performed using Autodock tools (ADT) following standard protocols (353) as follows:

a) The selected protein was cleared of any artefacts, water, unbound ions, secondary chains and ligands.
b) ADT were used to add polar hydrogens (for determining hydrogen bonds) and partial molecular charges (to ensure the correct electrostatic potential) to the model.

c) A gridbox was generated using ADT, and the model is converted to a *pdbqt* format using *Open Babel* toolkit.

For ligand assessments, including mycosporine-like amino acid (MAA) predictions, the Avogadro chemical utility platform (354) was used to create and optimize compound binding models by molecular dynamics simulation using universal force field (UFF) parameters (355) and the steepest descent algorithm (356). To assess MAA-Keap1 docking results, a comparison set of approximately 1,100 Brazilian natural products was assembled from two ZINC catalogs (357–359) for predictive contrast. All ligands were prepared for docking using the AutoDock Tools (ADT) script "prepare_ligand4.py" (for AutoDock Vina) and the UCSF Chimera visualization toolkit (360) (for DOCK6). Both Autodock Vina and DOCK6 were configured for high docking precision with DOCK6 selected for a flexible docking protocol with 2,000 orientations per ligand and 400 iterations for energy minimization, and the AutoDock Vina space search exhaustiveness was set to 20. The UCSF Chimera ViewDock utility was used to manually examine the docking results. Ligands were assessed first by the number of potential hydrogen bonds available for binding within the protein binding pocket, and all ligands forming less than 3 hydrogen bonds were rejected. Those selected were evaluated manually, and all ligands without potential binding to critical positions of the Keap1-Nrf2 docking pocket were discarded. Final evaluation of "viable" docking ligands, was performed by assessing the combined docking scores calculated as 2 x Vina docking energy + DOCK6 binding energy + DOCK6 docking score. Each variable in the equation was scaled by subtracting the mean value and dividing by the standard deviation.

2.5 Results

2.5.1 Distant homology search pipeline (DHSP)

Distant homology searches for animal and fungal homologs of human Keap1 and Nrf2 proteins were conducted using a custom built Distant Homology Search Pipeline (DHSP). DHSP is a semi-automatic tool for high sensitivity, high precision searching of distant homology. It was developed in Python and runs in a LINUX command line environment. The pipeline (Figure 2.1) uses psi-BLAST against the NCBI NR database for close homology searching to generate HMM models from best psi-BLAST results. It uses a HMMER search for high sensitivity distant homology detection. To ensure a low number of false positive hits, it performs searches using multiple HMM models and the Smith-Waterman algorithm to align potential homologs against the original sequence, filtering out those that fail to align properly and match multiple HMM models.

DHSP performs the following steps sequentially (pipeline parameters can be configured but the listed ones were used in this analysis):

a) The selected "target" protein and several manually selected known homologs of the "target" are used as inputs.
b) Psi-BLAST with three iterations is used to find close homology for each input sequence.
c) For every input, the highest scoring 100 homologs with e-values below $1.0e^{-10}$ and coverage of at least 70 % are aligned with Clustal Omega using default parameters.
d) *hmmbuild* tool with default parameters is used to generate HMM models of selected input sequences.
e) *hmmsearch* tool is used to search each of the custom databases with a HMM model of each input. Sequences with *hmmsearch* e-value cutoff of 1.0 or lower are accepted for further refinement.
f) EMBOSS package *water* implementation of Smith-Waterman algorithm is used to align all potential homologs to the original protein sequence, using the BLOSUM62 scoring matrix, gap-opening penalty 10.0 and gap-extension penalty 0.5. Sequences with a Similarity Value (as reported by EMBOSS package *water* implementation of Smith-Waterman and calculated as the sum of aligned identical amino acids and aligned highly conserved amino acids, divided by length of alignment) under 35 % are rejected.

g) Potential homologs detected by fewer than 40 % of HMM models (where one model is generated for the input sequence and each of its manually selected close homologs, see *step a*) are rejected.

To facilitate comparison with other tools, DHSP runs parallel searches using BLAST, psi-BLAST, HMMER, iterative HMMER and HHblits and compares the results from these tools. Results of the DHPS combined HMMER search and tool comparisons are grouped by NCBI taxonomy and are displayed graphically using the Taxonomy Landscape Mapper, described in section 2.5.2, and in pseudo-FASTA format designed for easy data post-processing by the TLM. Pseudo-FASTA is identical to standard FASTA format with addition of e-value within the sequence header in __eV{NUM#}Ve__ format (where NUM# is replaced by actual e-value of sequence similarity search and is used by TLM to assign scoring to generated taxonomical representation). The DHSP code is freely available at https://github.com/rgacesa/DHSP.

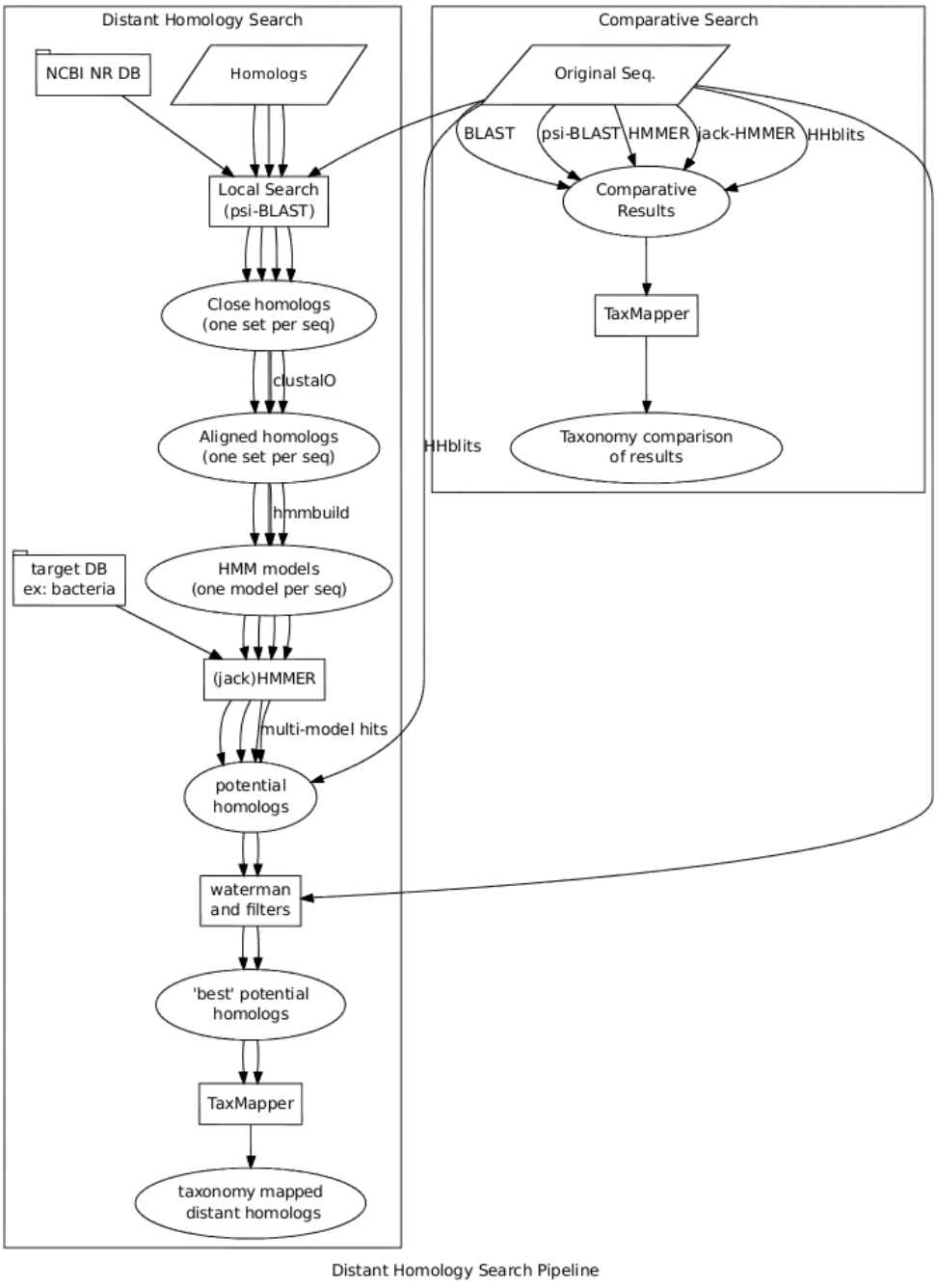


Figure 2.1 Workflow of the Distant Homology Search Pipeline (DHSP)

The DHSP input consists of the "original" sequence targeted for homology search and several manually selected (known and annotated) homologs. Each of the input sequences is BLAST searched against the NR database and the "best" (highest bit-score) homologs are multiple-aligned and converted into a HMM model. Each of these models is used to search target database(s) containing protein sequences with potential for homology. Results for multiple HMM models are compared and their validity is verified by a number of HMM models that produced the match and by alignment to the "original" input sequence. Parallel homology searches by one or more commonly used tools are produced for comparison, and all results are mapped to the NCBI taxonomy database using TLM.

2.5.2 Taxonomy Landscape Mapper (TLM)

Results of DHSP were mapped to NCBI taxonomy via the newly developed tool for homology search mapping. TLM links FASTA formatted sequences and the results of homology search tools (BLAST, psi-BLAST, HMMER, jack-HMMER and HHblits) to NCBI taxonomy (Figure 2.2). Written in Python, it runs as a LINUX command-line program with a large set of options for customizing input and output data and for results filtering. Its output includes taxonomical distribution of input sequences in text format and visual representation of the results (Figure 2.3). TLM code is freely available at https://github.com/rgacesa/TLM.

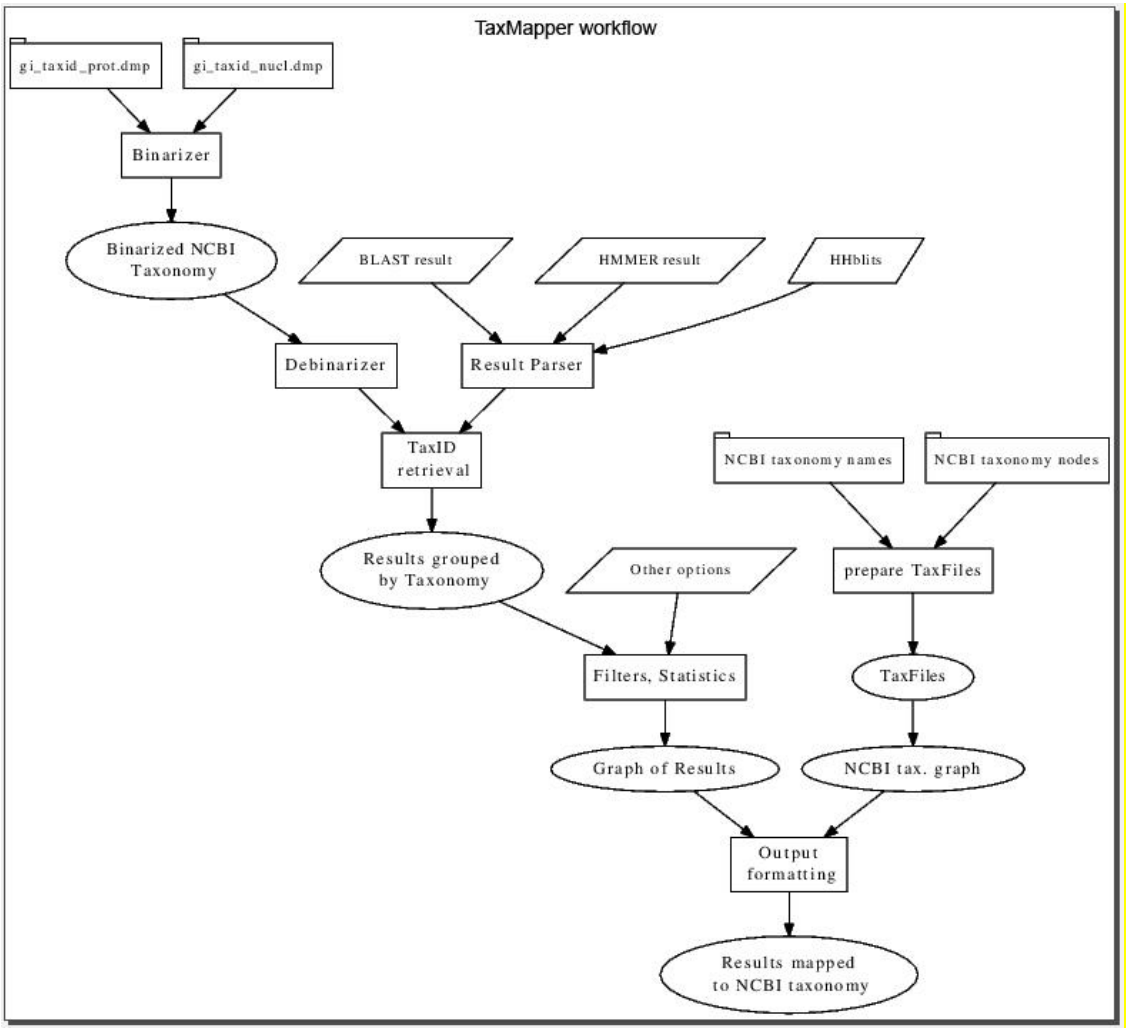


Figure 2.2 TLM workflow schematic

TLM extracts sequence identifiers and homology search scores from the results of HMMER, HHblits and BLAST searches. Sequences are linked to NCBI taxonomy via sequence identifier and assigned taxonomical classification. Data is subsequently paired with the entire NCBI taxonomy database represented as a directed graph. The TLM output consists of user-friendly tabular text output and raw text data formatted for input into the GraphViz drawing package.

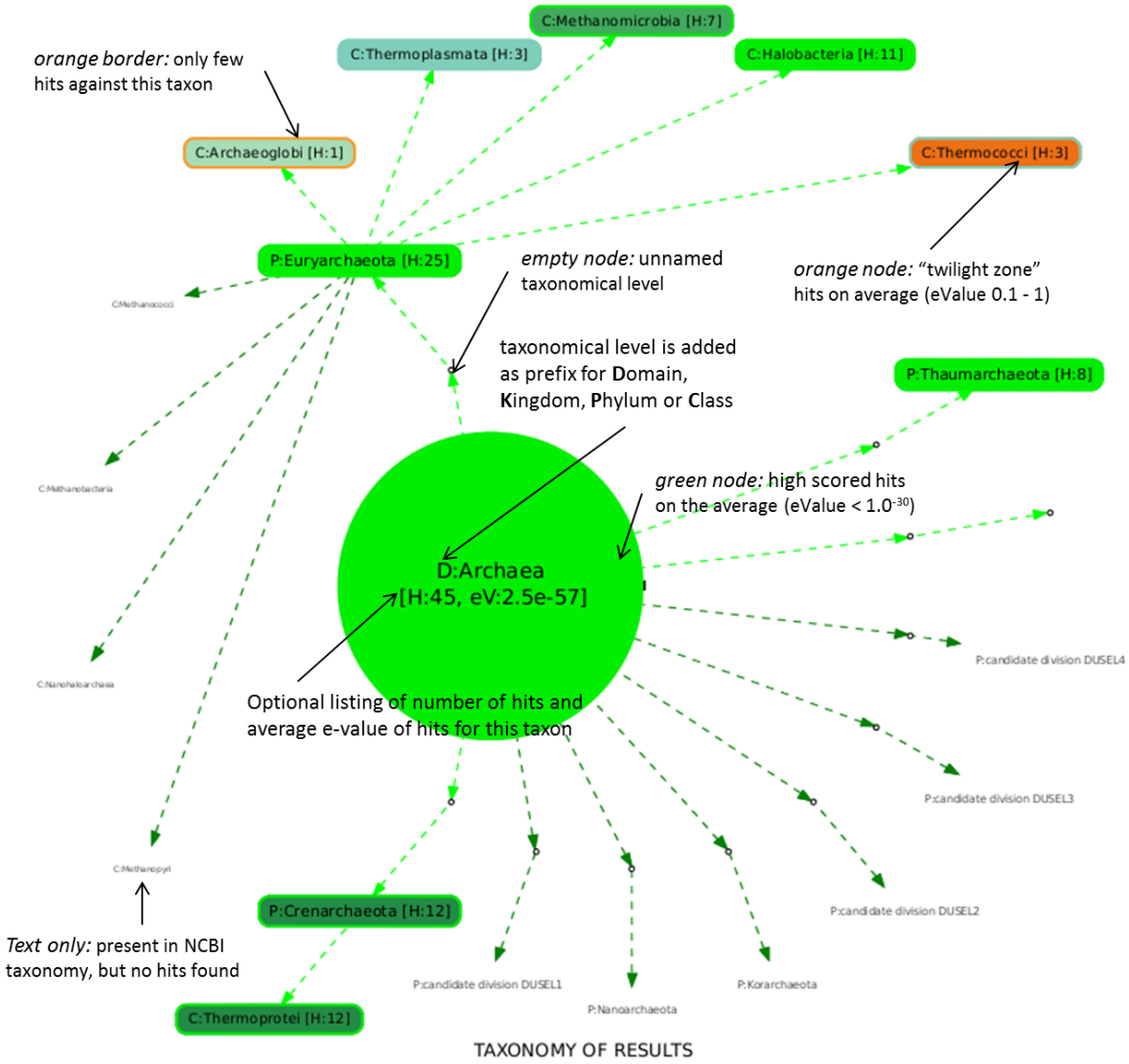


Figure 2.3 Taxonomy landscape mapper output

The Taxonomy landscape mapper generates output in raw text format suitable for direct conversion into vector graphics by GraphViz package. Taxonomy is represented as directed graph with nodes displayed according to taxonomical level and homology search score quality (green for very high scored hits, blue for high scored and yellow to red for "twilight zone" hits). Edges are coloured depending on number of detected homologs in relation to total number of detected homologs.

2.5.3 Data mining of microbial protein databases

Databases of archaea, bacteria, fungi and plant proteins were analysed for distant homology to human Keap1 and Nrf2 proteins using the newly developed software tools we named the Distant Homology Search Pipeline (DHSP) and the Taxonomy Landscape Mapper (TLM). All databases except for archaea were found to contain high numbers of Keap1 homologs (Appendix A-2). In contrast, close homologs to human Nrf2 domain Neh1 – Neh6 sequences were detected only in the database of fungal proteins (Appendices A-2 and A-3), primarily in Ascomycetes belonging to the Class *Sordariomycetes*, many of which are insect and plant pathogens (Table 2.1).

Table 2.1 Keap1 and Nrf2 protein scoring of sequence homology in fungal genomes.

The table shows species selected based upon the prediction score of detected homologs to Keap1 and Nrf2 proteins, all having the presence of genes for mycosporine-like amino acid (MAA) biosynthesis. Nrf2 prediction scores are the sum of numbers of detected homologs for Nrf2 domain Neh1 – Neh6 conserved sequences. The Keap1 prediction scores are the sum of Keap1 conserved kelch1 – kelch6 and BTB domain sequences.

Species (Common name)	**Prediction score**	
	Nrf2	**Keap1**
Cordyceps militaris (Scarlet caterpillar club fungus)	8	46
Beauveria bassiana (White muscardine fungus)	7	43
Fusarium graminearum (Wheat head blight fungus)	10	66
Colletotrichum gloeosporioides (Post-harvest fruit rot fungus)	8	79
Magnaporthe oryzae (Rice blast fungus)	15	79
Fusarium pseudograminearum (Wheat crown rot fungus)	10	60
Verticillium dahlia (Verticillium wilt fungus)	10	40
Colletotrichum higginsianum (Crucifer anthracnose fungus)	14	54
Metarhizium acridum (Green muscardine fungus)	13	38
Colletotrichum graminicola (Maze anthracnose fungus)	15	73

2.5.4 Phylogenetic reconstruction of Keap1-Nrf2 homologies

In order to examine evolution of the Keap1-Nrf2 pathway, we performed phylogenetic reconstruction using predicted Keap1 and Nrf2 fungal homologs, as well as a selection of known homologs from key animal species. These animals comprise invertebrate and vertebrate species commonly used as model organisms in biology that include three species from the genus *Caenorhabditis* and the fruit fly *Drosophila melanogaster* (325,329). The sponge *Amphimedon queenslandica* and coral *Acropora digitifera* were chosen as examples of early metazoan taxa (phyla Porifera and Cnidaria). Vertebrate sequences were chosen from common model organisms (Rat, Mouse, Zebrafish and frog *Xenopus laevis*) and their close relatives. *Platypus* was chosen as an example of early mammals and the lancet *Branchiostoma floridae* was selected to represent an ancient vertebrate animal. All amino acid sequences are listed in Appendix A-5. Bayesian reconstruction of both Keap1 phylogeny (Figure 2.4) and Nrf2 phylogeny (Figure 2.5) were consistent with conventional species evolution, with the expected grouping of major vertebrate taxa and a clear split between the vertebrates and invertebrates. Fungal Keap1 and Nrf2 sequences, however, were both highly divergent and could be grouped into 3 clades. Surprisingly, phylogenetic reconstruction of Nrf2 homologs placed all three *Caenorhabditis* proteins within the fungal groupings (Figure 2.5), instead of residing with other invertebrates as would be expected.

Figure 2.4 Bayesian phylogenetic reconstruction of Keap1 evolution.

The Bayesian phylogenetic tree was constructed from 16 vertebrate Keap1 homologs, 7 invertebrate homologs and 42 fungal homologs. Fungi are coloured brown, invertebrates blue and vertebrates green. Bayesian posterior probabilities are displayed for taxonomical splits with posterior probabilities above 0.5, and splits with lower posterior probabilities have been collapsed. Invertebrate taxa where Nrf2 has been experimentally confirmed are marked with a blue star.

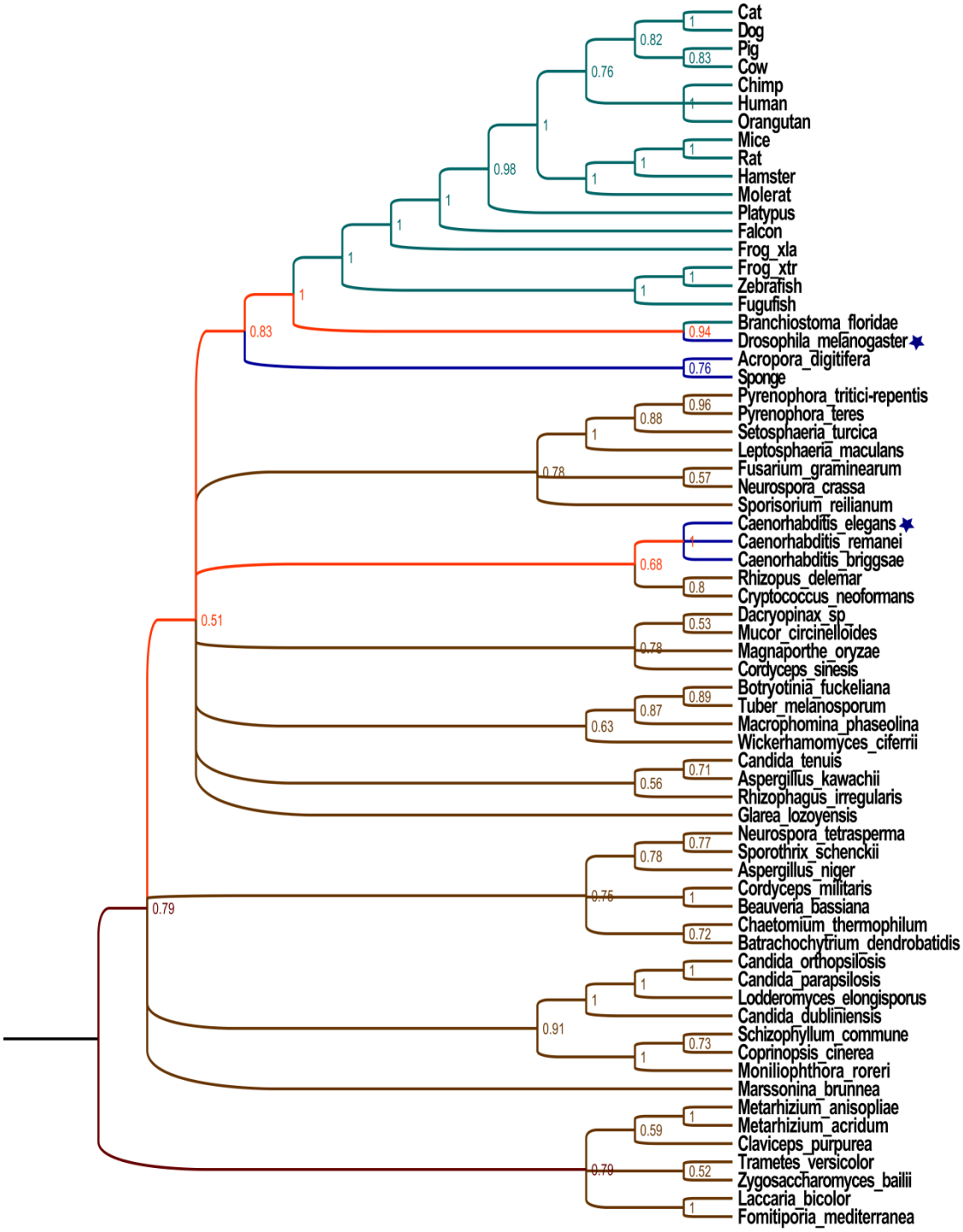


Figure 2.5 Bayesian phylogenetic reconstruction of Nrf2 evolution.

The Bayesian phylogenetic tree was constructed from 16 vertebrate Nrf2 homologs, 7 invertebrate homologs and 42 fungal homologs. The fungi are coloured brown, invertebrates blue, vertebrates green and polyphyletic branches are coloured red. Bayesian posterior probabilities are displayed for each taxonomical split with posterior probability above 0.5, and splits with lower posterior probabilities have been collapsed. Invertebrate taxa where Nrf2 has been experimentally confirmed are marked with a blue star.

2.5.5 Protein modelling and virtual screening of Nrf2 activation

Fungal genomes typically encode enzymes that express the biosynthesis of mycosporines and the related mycosporine-like amino acid (MAA) family of UV-protective and antioxidant metabolites (64,234). In order to assess whether MAAs have potential to initiate a protective response through the Keap1-Nrf2 pathway, we performed a virtual screen of approximately 1,100 diverse natural products including 20 MAAs. Of the ligands tested, 75 met the criteria for potential inhibitors of the Keap1-Nrf2 interaction. These criteria were determined by the docking position of the ligand within the Keap1-Nrf2 interaction pocket, the potential to form hydrogen bonds with Keap1, and importantly the docking score. Out of the 75 compounds (Table 2.2), 25 are known to be Nrf2 activators, while another 11 compounds are known antioxidants but not reported previously to activate Nrf2. These 11 compounds included 3 MAAs (mycosporine-glycine-valine, mycosporine-glycine and porphyra-334). Examples for the binding of betanidin and porphyra-334 within the docking region of Keap1 are shown in Figure 2.6.

Table 2.2 Virtual screening results.

Listed compounds are deemed "viable" according to their docking score, number of potential hydrogen bonds within the Keap1-Nrf2 binding pocket and by manual inspection of the docking profile. The table is sorted according to the biological function of docked ligands. Entries are given only for compounds with assigned structures; a complete list with unknown structure assignments is given in Appendix A-7.

Compound	Score	Structure Assignment	Biological function
ZINC49048037	0.75	AGN-PC-07CJ71	acetylcholinesterase inhibitor (361)
ZINC15120547	-2.66	Crassinervic acid	antifungal (362)
ZINC00622123	0.77	Griseofulvin	antifungal (363)
ZINC13411177	0.02	similar to Strictifolione	antifungal (364)
ZINC14447808	1.76	AGN-PC-077JEH	antifungal (365)
ZINC40973915	-9.01	similar to Ixoside	antioxidant (366)
ZINC31157290	-2.60	Secoxyloganin	antioxidant (367)
ZINC05998957	-2.17	Lirioresinol A	antioxidant (368)
ZINC15119278	-1.04	similar to Yatein	antioxidant (369)
ZINC00898006	-0.15	Rubrofusarin	antioxidant (370)
ZINC02563652	-0.04	Alloisoimperatorin	antioxidant (371)
ZINC01580260	0.23	Cleomiscosin A	antioxidant (372)
ZINC69482380	1.63	similar to Maclurin	antioxidant (373)
ZINC06037073	-0.97	similar to Emodin	cytotoxic, anti-cancer (374)

ZINC84154280	-2.54	Geranyloxy-p-benzoic Acid	farnesoid X receptor agonist (375)
ZINC26490614	-2.69	Procyanidin B2	Nrf2 activator (376)
ZINC30726399	-9.93	Betanidin	Nrf2 activator (377)
ZINC69482045	-6.62	similar to Ursoloic acid	Nrf2 activator (378)
ZINC69481913	-6.40	similar to Ursoloic acid	Nrf2 activator (378)
ZINC17263588	-6.17	Chlorogenic acid	Nrf2 activator (379)
ZINC84154032	-5.75	similar to Morroniside	Nrf2 activator (380)
ZINC84153764	-4.32	similar to Morroniside	Nrf2 activator (380)
ZINC04102166	-4.28	Geniposidic acid	Nrf2 activator (381)
ZINC01714287	-3.40	Piperine	Nrf2 activator (382)
ZINC03870412	-3.06	Epigallocatechin gallate (EGCG)	Nrf2 activator (383)
ZINC00073693	-2.12	Pinocembrin	Nrf2 activator (384)
ZINC12428433	-1.84	Butein	Nrf2 activator (385)
ZINC71316232	-1.69	similar to Chlorogenic acid	Nrf2 activator (379)
ZINC01531693	-1.57	similar to Piperine	Nrf2 activator (382)
ZINC03872070	-1.52	Chrysine	Nrf2 activator (386)
ZINC00897734	-1.50	similar to Quercetin	Nrf2 activator (387)
ZINC00156701	-1.41	Naringenin	Nrf2 activator (388)
ZINC00113309	1.69	Fraxetin	Nrf2 activator (389)
ZINC01561070	-0.11	similar to Quercetin	Nrf2 activator (387)
ZINC14728348	0.14	similar to Quercetin	Nrf2 activator (387)
ZINC05733652	-1.36	Diosmetin	potential Nrf2 activator, antioxidant (390)
ZINC33832113	-1.73	similar to Phlorizin	potential Nrf2 activator (391)
ZINC69482290	-3.37	similar to Glucoerucin	potential Nrf2 activator (392)
ZINC05733537	-0.86	Ermanin, similar to Quercetin (Nrf2 activator)	potential Nrf2 activator (387)
ZINC84153966	-3.86	similar to Acetoside	potential Nrf2 activator (393)
ZINC13108875	-2.42	similar to Burchellin	potential pesticide (394)
Mycosporine glycine-valine	0.24	Mycosporine-like amino acid	UV-protectant, antioxidant (395)
Mycosporine glycine	4.21	Mycosporine-like amino acid	UV-protectant, antioxidant (395)
Porphyra 334	0.58	Mycosporine-like amino acid	UV-protectant, antioxidant (395)
ZINC15252691	-5.23	Gaudichaudianic acid	trypancide (396)

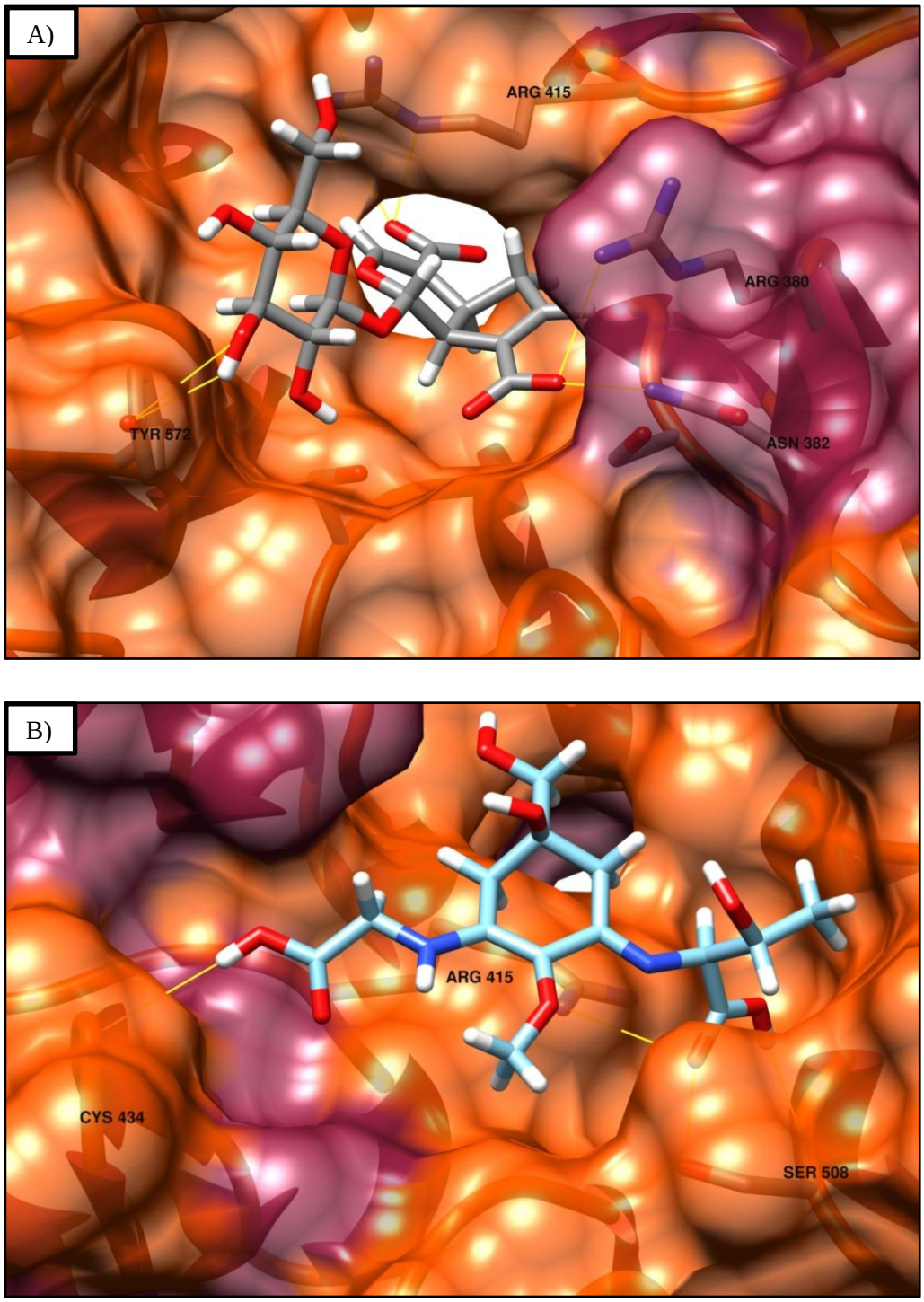


Figure 2.6 Betanidin (A) and porphyra-334 (B) substrate docking models.

The protein binding models depict the cross-section of only the molecular surface of the human Keap1 kelch-like repeats β-propeller docking pocket. Predicted hydrogen bonds between ligand and Keap1 are depicted in yellow and amino acid residues involved in the formation of bonds are labeled.

2.6 Discussion

The Keap1-Nrf2 pathway is a major regulator of antioxidant protection in mammalian cells, and is responsible for the transcription of over 200 cytoprotective genes encoded by the nuclear antioxidant response element (ARE). Given that the Keap1-Nrf2 pathway is so important for cytoprotection in mammals, it might be expected that homology is evolutionary preserved from simple progenitors. This is consistent with a homologous Keap1-Nrf2 pathway confirmed in *Drosophila melanogaster* (325) and identified in *Caenorhabditis elegans* (329). Our distant homology search has revealed, for the first time, that Keap1 and Nrf2 homologs are present in fungal taxa (Table 2.1 and Appendix A-2) and absent in bacteria, archaea and plants (Appendix A-2). The presence of both Keap1 and Nrf2 homologs in fungi, and that fungi are evolutionary closer to animals (397) than all other taxa examined, encouraged further investigation. Phylogenetic reconstruction of Keap1 homology in key species from vertebrate, invertebrate

and fungal taxa (

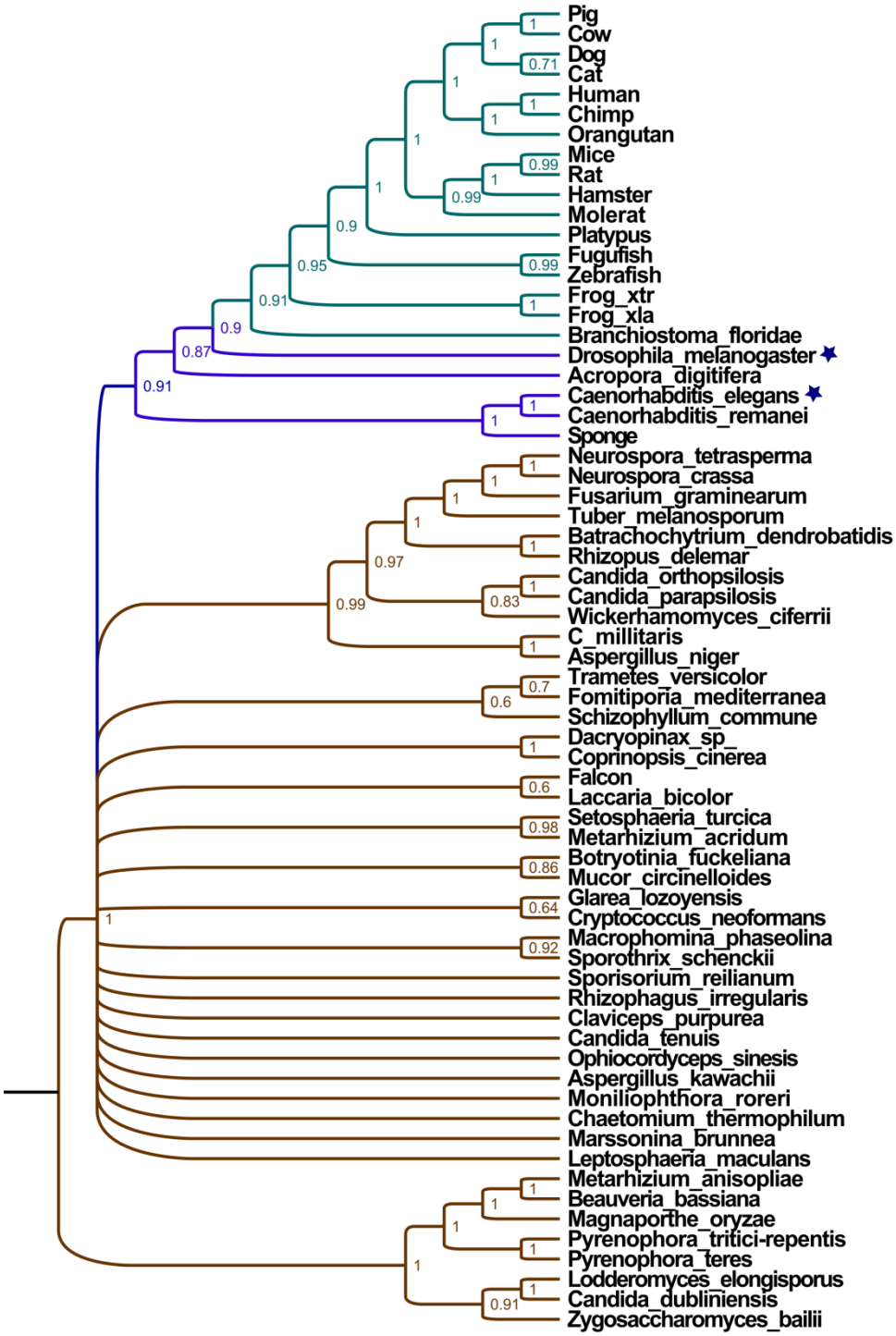


Figure 2.4) demonstrated that Keap1 fits as expected within the "Tree of Life" (398). However, phylogenetic reconstruction of Nrf2 (Figure 2.5) showed an unusual discrepancy in the placement of the genus *Caenorhabditis*. Genomic data mining confirms that *C. elegans* does encode proteins highly similar to Keap1 (Appendix A-8), consistent with its position in the

Keap1 phylogenetic tree (

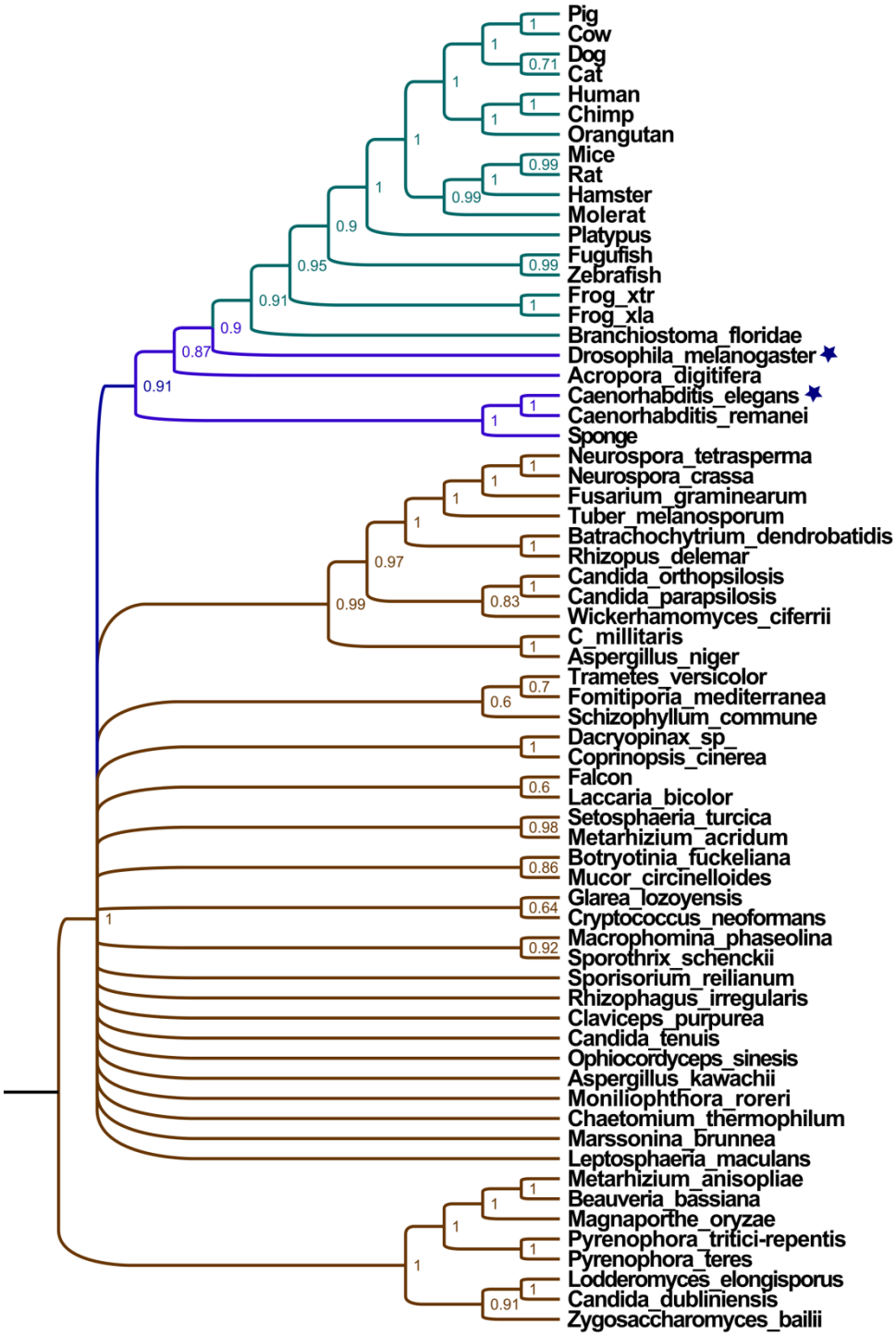


Figure 2.4). These Keap1-like proteins have not been implicated in inhibition the Nrf2 homolog of *C. elegans*, specifically the protein designated SKN-1 (329,399), and closer phylogenetic examination of this functional placement is warranted. Instead, SKN-1 activity is regulated by interaction with protein WDR-23 in a manner strikingly similar to Nrf2-Keap1-Cul3

ubiquitination and degradation (400). Interestingly, while WDR-23 has no significant sequence similarity to Keap1 (Appendix A-8), the 3D structure is remarkably similar to that of Keap1, with both proteins containing a beta-propeller superstructure. Unlike Keap1, however, the WDR-23 beta-propeller is formed from WD-40 protein motif repeats (see http://www.rcsb.org/pdb/explore/explore.do?structureId=4LG9) rather than kelch-like repeats (see http://www.rcsb.org/pdb/explore.do?structureId=2flu). Importantly, recent research has established that Nrf2 activity is controlled additionally by the human beta-transducin repeat-containing protein (β-TrCP) (167,168), which contains WD-40 protein motifs arranged into a beta-propeller superstructure (401). Comparison between *C. elegans* WDR-23 and the human β-TrCP protein (Appendix A-8) reveals significant similarity between these two WD-40 beta propeller proteins, indicating remarkably strong evolutionary continuity of function for Nrf2 control shared between worms and higher animals.

According to our predictions of Nrf2 homologs in various fungi (Figure 2.5 and Table 2.1) and previous findings of the Nrf2 homolog Yap1 in yeast (330), the Keap1-Nrf2 pathway has most certainly evolved after the eukarya separated from the prokarya, but prior to the fungal-metazoan split. Therefore, beta-propeller inhibitor proteins of Nrf2 may have evolved in early animals more than once by convergent evolution, which might explain the observed differences of Nrf2 activation in *C. elegans* and that of higher animals. Prior bioinformatics analysis of the sea anemone *Nematostella vectensis* has also detected homologs of Nrf2, Keap1 and two small binding Maf proteins required for Nrf2-ARE gene promotion (402). Our analyses of coral (*Acropora digitifiera*) and sponge (*Amphimedon queenslandica*) genomes also revealed more than one small Maf homolog [data not shown], in addition to encoding both Keap1 and Nrf2 homologs. These results indicate that species within the genus *Caenorhabditis* may have lost SKN-1 regulation by a Keap1-like protein (possibly R12E2.1), but have retained a β-TrCP like inhibitor WDR-23. This loss may be due to a lack of evolutionary pressure associated with the soil-dwelling, often hypoxic, lifestyle of these animals (403). Further investigation by inclusion of additional deep-branching taxa is required; nevertheless, our phylogenetic analyses show clearly that the Keap1-Nrf2 pathway predates the fungal-metazoan divergence. Developing an evolutionary clock to determine if there is a correlation between the emergences of Nrf2 with the generation of an oxygen atmosphere on Earth is an avenue worthy of future research.

A virtual screening assay was performed to assess the potential for fungal metabolites to function as competitive inhibitors of Keap1-Nrf2 binding. Although there are no published data for the disruption of Keap1-Nrf2 binding by ixoside metabolites, our predictions (Table 2.2) match the high scoring antioxidants betanidin, chlorogenic acid and compounds similar to

ursolic acid, which are known activators of Nrf2 (377,378,404). We found also that the MAAs, mycosporine-glycine, mycosporine-glycine-valine and porphyra-334, may serve as viable docking ligands based on their docking score, docking profile and potential to form critical hydrogen bonds within the Keap1-Nrf2 docking pocket. All three MAAs, often expressed in high cellular concentrations, are widely accepted to be UV-inducible sunscreen protectants (64) and mycosporine-glycine and porphyra-334 are reported to have antioxidant properties (250,255). While docking scores of MAAs are on the average higher than many of the other viable ligands, implying potentially lower binding affinity, these compounds are predicted to form several hydrogen bonds with Keap1 binding pocket and have passed manual inspection of binding poses. Additional research will determine if these compounds may cause disruption of Keap1-Nrf2 binding to activate the transcription of nuclear ARE cytoprotective genes. Finding microorganisms with Keap1-Nrf2 homology offer an early evolutionary model for the adaptive signalling of the Keap1-Nrf2 pathway, as well as providing an endogenous source of stress-inducible metabolites having potential to activate the nuclear ARE for therapeutic consideration.

2.7 Conclusions

Data mining of microbial protein databases has revealed distant homology to Keap1 and Nrf2 proteins in fungi, especially amongst taxa of Phylum: Ascomycota / Class: Sordariomycetes. Phylogenetic reconstruction of Keap1-Nrf2 pathway shows that the pathway evolved prior to the fungal-metazoan divergence. Unexpectedly, the Nrf2 evolutionary tree shows mismatch for genus *Caenorhabditis* within the expected taxonomic model, potentially from sequence degeneration of Nrf2 or lack of evolutionary pressure possibly due to the soil-dwelling lifestyle of these worms. Lastly, virtual screening for competitive inhibition of Keap1-Nrf2 binding predicts the potential for Nrf2 activation by UV-protective mycosporine-like amino acids.

Chapter 3: Rising levels of atmospheric oxygen and evolution of Nrf2

The content of this chapter was published as:

Gacesa R., Dunlap W.C., Barlow D.J., Laskowski R.A. and Long P.F. (2016). **Rising levels of atmospheric oxygen and evolution of Nrf2**. Scientific Reports 6, Article number 27740

R.G. designed the study, assembled the data, carried out phylogenetic analysis, interpreted the results and drafted the manuscript.

Detailed bioinformatics methodology (sections 3.4.1 to 0) included in this chapter is published as part of *Supplementary Data File 1: Bioinformatics methodology*, and is also available online at http://www.nature.com/articles/srep27740#supplementary-information.

In-detail list of sequences used in phylogenetic reconstruction is included in the electronic format on the *Appendix Disk* as *Appendix B* and is also available online at http://www.nature.com/articles/srep27740#supplementary-information.

3.1 Foreword to Chapter 3

The Chapter 2 of this thesis described the results of phylogenetic reconstruction of evolution of genes encoding Keap1 and Nrf2 proteins in eukaryotes, and demonstrated that all tested animal and fungal genomes contained homologs to vertebrate genes encoding Keap1 and Nrf2. Research presented in this chapter builds upon the conclusions of Chapter 2, and establishes the results of application of evolutionary clock hypothesis to the phylogenetic tree of Nrf2 protein. In addition, the evolutionary time-frame of Nrf2 is correlated with the atmospheric oxygen levels over geological time to assess the impact of the atmospheric oxygen levels on the evolution of the Nrf2 pathway.

3.2 Abstract

In mammals, the master transcription regulator of antioxidant defences is provided by the Nrf2 protein. Phylogenetic analyses of Nrf2 sequences are used here to derive a molecular clock that manifests persuasive evidence that Nrf2 orthologues emerged, and then diverged, at two time points that correlate with well-established geochemical and palaeobiological chronologies during progression of the 'Great Oxygenation Event'. We demonstrate that orthologues of Nrf2 first appeared in fungi around 1.5 Ga during the Paleoproterozoic when photosynthetic oxygen was being absorbed into the oceans. A subsequent significant divergence in Nrf2 is seen during the split between fungi and the Metazoa approximately 1.0 – 1.2 Ga, at a time when oceanic ventilation released free oxygen to the atmosphere, but with most being absorbed by methane oxidation and oxidative weathering of land surfaces until approximately 800 Ma. Atmospheric oxygen levels thereafter accumulated giving rise to metazoan success known as the Cambrian explosion commencing at ~541 Ma. Atmospheric O_2 levels then rose in the mid Paleozoic (359-252 Ma), and Nrf2 diverged once again at the division between mammals and non-mammalian vertebrates during the Permian-Triassic boundary (~252 Ma). Understanding Nrf2 evolution as an effective antioxidant response may have repercussions for improved human health.

3.3 Introduction

The 'Great Oxygenation Event' (GOE), at 2.45-1.85 Ga is recognised as the most geologically critical environmental change impacting the history of life on Earth (405). Oxygen-producing photosynthetic cyanobacteria appeared much earlier, preceding the increase of atmospheric oxygen marked by the onset of the GOE (406), but this oxygen was removed from the atmosphere by rapid oxidation of reduced minerals, precipitating especially vast deposits of ferric oxide from the oxidation of dissolved oceanic ferrous iron. Only after this mineral oxygen sink approached saturation, a process colloquially referred to as the 'Rusting of the Earth', did atmospheric oxygen increase at the advent of the GOE, giving a time-lag from the origin of oxygen-producing photosynthetic cyanobacteria that seems to have lasted ~1 Ga (405). The GOE provided biologically useable molecular oxygen necessary for aerobic respiration, a decidedly more efficient energy-generating process than pre-existing metabolic pathways, thus setting the stage for an evolutionary transition to the aerobe-dominated biota that continues to this day.

An important problem key to the success of the history of aerobic life on Earth is how cellular processes co-adapted to overcome the metabolic toxicity that results from use of highly reactive molecular oxygen. In aerobic respiration, enzyme catalysed four-electron reduction of oxygen is considered to be a relatively safe process producing water at the terminal end of the mitochondrial electron transport chain. The reductive environment of cells, however, provides ample opportunities for oxygen to undergo successive non-enzymatic univalent reduction, these processes being exacerbated by electrophilic xenobiotics and abiotic agents such as solar ultra-violet radiation. Oxidative stress is the net outcome of oxidative damage to biologically important molecules such as proteins, lipids, carbohydrates and nucleic acids caused by the generation of these reactive oxygen species (RS). To survive in such a reactive oxygen environment, living organisms produce or sequester a variety of water- and lipid-soluble antioxidant compounds such as vitamins C and E. Oxygen metabolising organisms additionally produce an arsenal of antioxidant enzymes that inactivate RS. Animal genomes often express over 200 antioxidant and xenobiotic detoxifying enzymes (407). The regulated induction and expression of these genes to protect against metabolically induced oxidative stress and electrophilic toxicity is co-ordinated by a small number of related nuclear transcription factors of the bZip/CNC family of proteins, the most important of these being the master regulator, nuclear factor erythroid 2-related factor 2 (Nrf2). The Kelch-like ECH-associated protein 1 (Keap1) forms an anchor complex with Nrf2. This complex dissociates in response to RS and toxic electrophiles, thereby releasing Nrf2 which then binds to the nuclear antioxidant response

element (ARE) and co-ordinates transcription of multiple antioxidant and detoxifying enzymes (217).

The domain architecture of Nrf2 is highly conserved across many diverse species of aerobic organisms. Our previous phylogenetic analyses clearly revealed that, whilst absent in bacteria, archaea and plants, the Keap1–Nrf2 pathway predates the fungal–metazoan divergence (408). Here we present a 'molecular clock' which estimates that the evolutionary origins of Nrf2 is allied to the timing of the global transition from anaerobic to aerobic conditions, and provides first demonstration of a metabolic adaptation in multiple eukaryotic ancestors having evolved a significant molecular response to the GOE.

3.4 Methods

The Nrf2 phylogenetic tree was constructed using BEAST version 2.3.0 (409) using a selection of Nrf2 homologs sourced from major metazoan and fungal phyla, and basic leucine zipper transcription factors from plant and cyanobacteria are utilised as outgroups. Sequences were aligned using T-Coffee Expresso (410) and T-Coffee Psi-Coffee (410) aligners and were evaluated using the T-Coffee TCS method to verify multiple alignment transitional consistencies (411). The phylogenetic tree was calibrated based on best paleontological estimates for the emergence of Eukaryota, the metazoan-fungal split and a set of animal phyla divides using compiled data from previous studies (282,397,412,413). In order to assess the robustness of phylogenetic reconstruction and selective pressures in the evolution of Nrf2 based on increasing oxidative stress, data were split into subgroups (Mammals, Amniotes, Tetrapods, Vertebrates, Deuterostomia, Bilateria, Eumetazoa and early Eukarya datasets) to examine protein and DNA sequence divergence using MEGA 6 (267). For each group, sequences were aligned using ClustalW (345), and alignments were analysed using the HyPhy test of codon selection and a codon-based Z test of selection for DNA sequences (414). Accordingly, Maximum Likelihood and Neighbour Joining Trees were constructed for each group, and tree topologies were compared to verify consistency of results. Tests confirmed the robustness of taxonomical grouping and codon-based selection tests within and between animal subgroups (data not shown). Multiple alignments with Nrf2-like DNA plant sequences were not of sufficient quality to perform Codon-based tests of selection.

3.4.1 Selection of sequences for phylogenetic reconstruction

Translated genomes of metazoan and fungi deposited in UniProt and NCBI RefSeq databases as of 01/06/2015 were data mined for homologs to human Nrf2 using HMMER (344) (HMM profiles generated for Nrf2 and Neh1 – Neh7 conserved sequences of Nrf2 using vertebrate Nrf2 sequences), psi-BLAST (343) and a previously developed Distant Homology Search Pipeline (DHSP (408)). If more than one homolog could be identified in a given genome, all potential homologs were investigated for Keap1 binding motifs DLG and ETGE and beta-TRCP binding motif DSGIS using pattern matching, with one mismatch and putative homolog selected based on the presence of DLG / ETGE motifs and HMMER e-values for Neh motifs. In the case of ambiguous results, pairwise BLAST alignment with human, mouse and *Drosophila* Nrf2 sequences were used to select putative homologs. DNA sequences were selected as coding DNA for Nrf2 protein homologs if available, and by BLAST searches against NCBI nucleotide databases if putative Nrf2 homologs lacked annotated coding sequences.

3.4.2 Reconstruction of dated phylogenetic tree

A dated phylogenetic tree was constructed using the BEAUTI/BEAST 2.3.0 framework (409), using the following 63 protein sequences from a set of major metazoan phyla. Plant and bacterial sequences were used as out-groups (see Appendix B-1 for list of sequences). Sequences were aligned using T-Coffee (415), M-Coffee (265), T-Coffee Expresso (410), Psi-Coffee (410), ClustalW (345), MUSCLE (416) and MAFFT (417) multiple alignment tools, with two independent runs for each tool. Each alignment was evaluated using T-Coffee TCS (411) for transitional consistency. Based on TCS scores, Expresso and Psi-Coffee were chosen as aligners of choice and three independent alignments were generated by each of these methods. Phylogenetic trees were constructed for each multiple alignment, using the following BEAST parameters:

- JTT evolutionary model (273), with Gamma site rates (Substitution rate, Proportion of invariant sites and Shape estimated during simulation, 4 gamma categories)
- Relaxed exponential clock model, with estimated rates and continuous rate variations along the tree
- Simulation was run for 100 000 000 MCMC generations

Following date ranges were used for calibration points (282,397,412):

- Bacteria-Eukarya divergence: ≈ 2200-4200 Ma (uniform prior probability; min 2200, max 4200 Ma; constrained as monophyletic outgroup)
- Bird-Reptile split: ≈ 255-300 Ma (gamma distributed prior probability; alpha 1.25, beta 10.0, offset 255.0 Ma)
- Eumetazoa – Metazoa divergence: ≈ 550-950 Ma (gamma distributed prior probability; alpha 1.25, beta 85.0, offset 550.0 Ma)
- Fungi – Animal divergence: ≈ 900-1500 Ma (normally distributed probability; mean 1200, sigma 100 Ma)
- Human – Chimpanzee split: ≈ 6 – 7 Ma (gamma distributed prior probability; alpha 1, beta 0.2, offset 6.0 Ma)
- Human – Mouse split: ≈ 69 – 110 Ma (gamma distributed prior probability; alpha 1.25, beta 8.0, offset 69.0 Ma)
- Plant – Animal split: ≈ 800 – 2000 Ma (normally distributed probability; mean 1400, sigma 200 Ma)
- Vertebrates – Invertebrates split: ≈ 500 – 600 Ma (gamma distributed prior probability; alpha 2, beta 15.0, offset 500.0 Ma)

Final trees were generated using *treeannotator (BEAST 2.3.0 package)* with *burnin value 0.25*, with other parameters left at default values. Trees were manually compared for consistency. The tree presented in the main article was generated using the *Figtree* tool with species from the same phylum collapsed for clarity, and posterior probabilities calculated as the mean between all BEAST runs. Comparison of trees found that all splits were highly consistent, even within clades with low posterior probability support.

3.4.3 Selective pressure analysis

Evolutionary selective pressure analysis was conducted using HyPhy test of codon selection and a codon-based Z test of selection (414) for DNA sequences by tools integrated into MEGA 6.0 toolkit (267). DNA sequences used for these tests are listed in Appendix B-2.

3.4.4 Data robustness analysis

In order to confirm the robustness of the data, DNA and protein sequences (Appendix B) were divided into the following subgroups, by mapping the NCBI sequence identified to the NCBI taxonomy database:

- Mammals
- Reptiles and Birds
- Land dwelling vertebrates
- All vertebrates
- Bilaterian animals
- Metazoa

And each group was further analysed using MEGA 6.0 by following protocol:

1. Sequences in the group were aligned using ClustalW and MUSCLE (using default parameters)
2. Maximum likelihood models were analysed using the MEGA Maximum likelihood (ML) model selection tool (model with lowest BIC and AICc scores were picked as models for choice)
3. Phylogenetic trees were reconstructed for each alignment using Neighbour joining and Maximum likelihood methods, using total deletion method and partial deletion method with cut-off of 95 % position coverage.
4. HyPhy test of codon selection and a codon-based Z test of selection were performed on the group.

In addition, multiple alignments used for dated tree reconstruction were also analysed using MrBayes, version 3.2 (338), using the following parameters for reconstruction of an undated phylogenetic tree:

- Prior for amino acid model set to mixed (*aamodelpr=mixed*), with gamma model invariant sites
- 10 000 000 MCMC generations, with 8 parallel chains and 4 runs
- Other parameters left at default values

Results of all tests were compared, tree topologies and dN-dS values were found to have high consistency between and within groups, with ClustalW alignments and partial deletion methods

generating results with high agreement to BEAST and MrBayes reconstructions. MUSCLE alignments and total deletion methods generated lower bootstrap values.

3.5 Results and Discussion

In order to reconstruct the evolutionary life history of Nrf2 in response to mounting oxidative stress, a Bayesian phylogenetic analysis of Nrf2 sequences retrieved from the genome sequences of many diverse taxa was performed together with a prediction of evolutionary pressure, calculated as ratio of synonymous to non-synonymous nucleotide base substitution rates. The results are presented as a phylogenetic tree which was converted to a "molecular clock" using widely accepted paleontological estimates for known splits between major animal phyla. The molecular clock was calibrated based on best paleontological estimates for the divergence of major phyla using compiled data from previous studies, reflecting the very recent hypothesis of Hedges et al. (2015) that speciation is independent of adaptation (282,412). The resulting phylogenetic reconstruction was mapped against the changing level of atmospheric oxygen over geological time – with the Phanerozoic oxygen levels taken as a composite of the data afforded from the GEOCARBSULF model of Berner (418–420), the glaciation-linked oxygen rise models of Harada et al. (421) and a compilation of other data (405,422–425). It should be noted that the oxygen curve presented in Figure 3.1 is based on "best estimates" and should thus be considered semi-quantitative. While Phanerozoic oxygen trends are well established (418–420), with moderate error margin (419), there is still a level of uncertainty over Proterozoic oxygen estimates. Specifically, the estimated date of origin of photosynthesis ranges from 2,400 to 3,000 Ma (425–427) and the exact oxygen levels over the majority of the Proterozoic era are subject to controversy (428–430), as are oxygen level dynamics during the Ediacaran era (421,424,431). Thus, while future research might lead to fine tuning of the oxygen level data, the pattern of change presented in Figure 3.1 is considered reliable as regards the *major* trends in oxygen change over geological time.

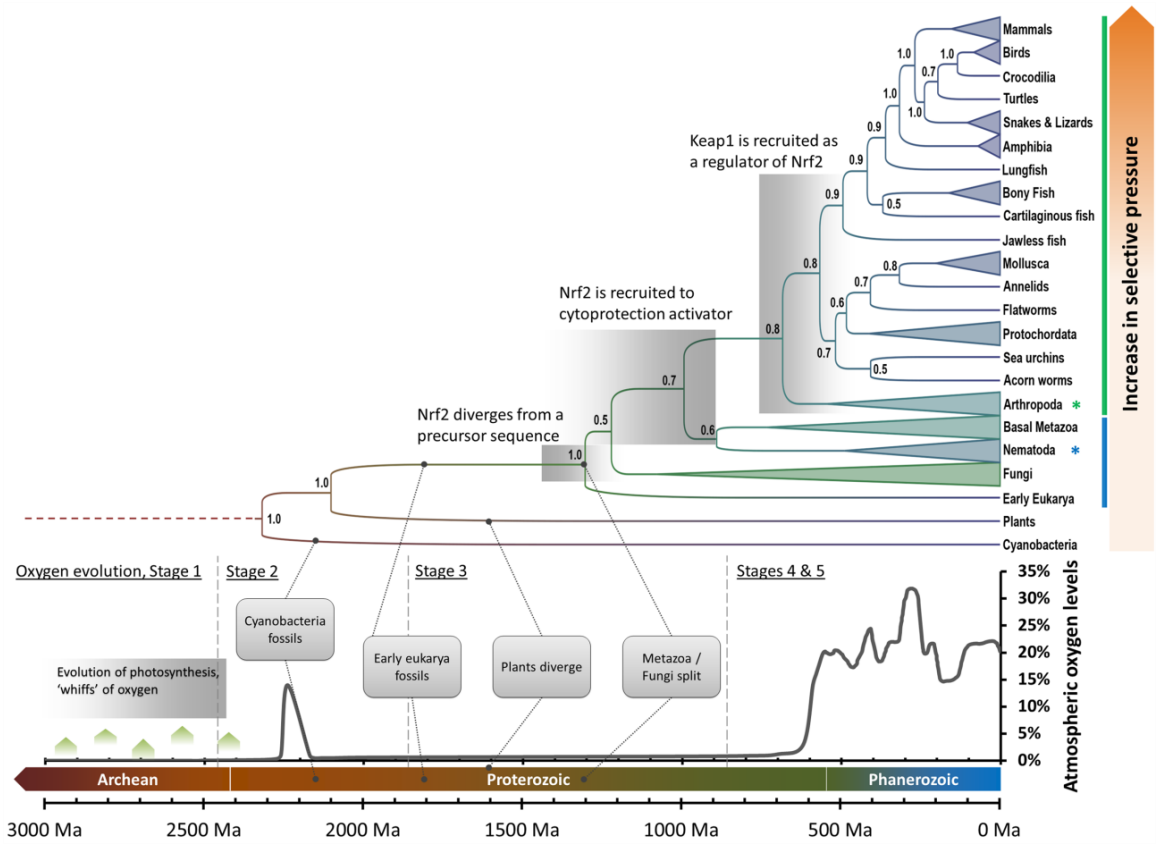


Figure 3.1 Nrf2 phylogenetic tree relative to atmospheric oxygen levels

The chart presents the Nrf2 phylogenetic tree relative to atmospheric oxygen levels during the latter period of Earth's history. The chart presents the traditional "5-stage model" of oxygen evolution constructed from compiled data (405,418–425), with the trend line representing a "best guess" model; Stage 1 represents a period when the atmosphere and oceans were largely anoxic; Stage 2 commences the 'Great Oxygenation Event'; Stage 3 is the period during which atmospheric oxygen levels remained low due to continued absorption by the oceans and oxidative weathering of the terrestrial crust; Stage 4 is the period after saturation of global oxygen buffers, during which oxygen levels rise towards present (Stage 5) atmospheric levels (PAL). The Earth timeline and major geological periods (432) are compiled and coloured by age. Eukarya and cyanobacteria appearances are noted according to first confirmed fossil evidence (427). Proposed time frames are shown for major Nrf2 divergence and recruitment events. Taxa known or predicted to contain the Keap1-Nrf2 signaling pathway are denoted by the vertical green bar, while taxa containing Nrf2 only (without Keap1) are denoted in blue. Invertebrates with an experimentally validated Nrf2 system are marked with a star (*). Evolutionary pressure increases towards more recently evolved phyla as schematically shown by an increasing orange hue in the selective pressure bar (decrease in dN-dS test statistic and decrease in p-value for null hypothesis of neutral evolution).

The results presented in Figure 3.1 allow inference of Nrf2 emergence and sequence diversification as speciation occurred and oxidative stress increased due to changes in atmospheric oxygen. These data would strongly suggest, therefore, that Nrf2 first appeared having evolved from an early eukaryotic peptide that contained a bZip/CNC domain sequence in Stage 3 of atmospheric oxygenation during the mid-Proterozoic when oxygen was released into the atmosphere but was rapidly absorbed into the Earth's ocean sediments and terrestrial crust (423). The divergence of cyanobacterial Nrf2-like sequences, which we use as an out-group in our evolutionary tree (Figure 3.1), differ in evolutionary time from the expected eukaryote-plant divergence (1500 Ma) (412), placing plant Nrf2-like sequences closer to cyanobacterial sequences rather than those of early eukaryotes. This significant difference in the bZip/CNC domain architecture of plants is consistent with a lack of nuclear Nrf2-like activation in the response of plants to oxidative stress (433) and absence of detectable homology to the Keap1-Nrf2 pathway in plant genomes (408). Perhaps Nrf2-like sequences in plants might be explained by horizontal transfer during early endosymbiosis, assuming such sequences were inherited from a cyanobacterial precursor of the plant chloroplast (434). Interestingly, the predicted oxygen level spike during Stage 2 of atmospheric oxygenation during the mid-Paleoproterozoic period does not seem associated with Nrf2 evolution (as evidenced by the lack of Nrf2 homology in cyanobacteria and plants (408)) and instead points to an Nrf2-like mechanism as a Metazoan adaptation.

Evolutionary pressure determined by the Codon-based Z-test of selection on both nucleotide and amino acid sequences (414) (Table 3.1) reveals strong purifying selection of Nrf2 sequences for all bilateral animals (all p-values $\leq 10^{-3}$) with the exception of nematode worms. Cnidaria and other basal metazoans display limited evidence for negative selection (p-values of all tests fall between 0.01 and 0.10). Nematodes and non-metazoan Nrf2 sequences exhibit no significant evidence for selective pressure (p-values of tests are > 0.10, as displayed in Table 3.1). Regulation of the Nrf2 antioxidant response exists in simple invertebrates as demonstrated empirically for *Caenorhabditis elegans.* The Nrf2 homolog SKN-1 in *C. elegans*, although serving a similar function, has significant differences in structure and regulatory pathways (324,329), lacking also a regulatory Keap1 interaction that is present in *Drosophila melanogaster* (325). Notably, the SKN-1 sequence of *C. elegans* is closer to the homolog sequences of basal metazoans such as cnidarians, indicating that recruitment occurred prior to the metazoan radiation of the Cambrian Explosion. This time frame matches the transition from Stage 3 and the start of Stage 4 of atmosphere oxygenation during which oxygen absorbing buffers in the Earth's oceans and crust were reaching saturation and atmospheric oxygen levels began to rise (405). This rise in atmospheric and ocean oxygen levels led to an increase in

aerobic metabolic stress causing evolutionary pressure towards the expansion of antioxidant response systems in animals. Tests for selective pressure indicate that Nrf2 sequences of basal metazoans were under limiting selective pressure, with averaged p-values for evolution neutrality > 1 (testing H_0: dN = dS, Table 3.1). According to empirical evidence gained from *Drosophila melanogaster* (325,435), genomic Keap1 recruitment occurred in early invertebrates preceding the divergence of the Class Insecta after the Cambrian Explosion. This time frame coincides with rising levels of atmospheric O_2 during Stage 4 of the Earth's oxygenation and matches the increased evolutionary pressure (measured by Codon-based Z-test of selection, Table 3.1) detected in Nrf2 sequences from taxa of the early Bilateria. Together, these lines of evidence suggest that rising levels of oxygen led to recruitment of Keap1 for enhanced regulation of Nrf2 for the transcription of cytoprotective genes in the response of animals to oxidative stress.

Table 3.1 Codon-based Z test of selection matrix

Table displays the Codon-based Z test of selection matrix performed on 42 DNA sequences of Nrf2 homologs from major eukaryotic phyla with cyanobacterial sequence used as outgroup (plant outgroup sequences could not be aligned with the dataset). Analyses were conducted using the Nei-Gojobori method, and results are grouped by major eukaryotic phyla with the phylum dN-dS value calculated as the mean of group members. All positions with less than 95% site coverage were eliminated. There were a total of 352 positions in the final dataset and fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. Evolutionary analyses were conducted in MEGA6. Table shows dN-dS and the p-value for null hypothesis of strict neutrality (dN=dS) for each pair of phyla. Phyla likely to have lower selective pressure compared to vertebrates (median P-value > 10^{-3}) are highlighted in yellow. Phyla without selective pressure (based on a p-value of 0.05 as the significance threshold) are highlighted in orange.

Phylum	Birds		Reptiles		Amphibia		Fish		Mollusca		Acorn worm		Echinoderm		Proto-chordata		Arthropods		Cnidaria		Nematoda		Early Eukarya		Fungi		Bacteria		dN-dS	pValue
	dN-dS	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*dN-dS*	*pv*	*mean*	*median*
Mammals	-15.40	0.0	-15.95	0.00	-18.15	0.00	-16.65	0.00	-7.73	0.00	-11.46	0.00	-6.99	0.00	-7.90	0.00	-7.40	0.00	-5.10	0.02	-2.16	0.09	-3.07	0.02	-2.31	0.17	0.20	0.28	-8.88	0.00
Birds			-10.54	0.00	-14.80	0.00	-13.64	0.00	-7.90	0.00	-9.90	0.00	-7.50	0.00	-7.20	0.00	-7.05	0.00	-5.38	0.00	-0.73	0.48	-3.10	0.01	-2.09	0.18	-0.12	0.68	-7.56	0.00
Reptiles					-14.87	0.00	-13.60	0.00	-6.77	0.00	-10.50	0.00	-6.29	0.00	-6.87	0.00	-7.14	0.00	-3.88	0.00	-0.92	0.40	-4.02	0.00	-2.28	0.08	-0.60	0.45	-7.66	0.00
Amphibia							-13.90	0.00	-6.64	0.00	-9.96	0.00	-6.02	0.00	-5.08	0.00	-7.53	0.00	-3.91	0.05	-0.85	0.45	-3.18	0.00	-1.58	0.17	0.53	0.60	-7.36	0.00
Fish									-7.69	0.00	-8.51	0.00	-6.65	0.00	-6.34	0.00	-6.85	0.00	-4.51	0.00	-1.49	0.20	-4.07	0.06	-1.79	0.24	-0.49	0.47	-8.06	0.00
Mollusca											-6.52	0.00	-8.76	0.00	-5.91	0.00	-7.92	0.00	-2.65	0.36	1.53	0.44	0.00	0.00	0.17	0.54	-0.81	0.42	-4.83	0.00
Acorn worm													-6.52	0.00	-6.09	0.00	-5.93	0.00	-3.37	0.01	1.13	0.26	-6.25	0.00	-0.72	0.56	0.25	0.80	-6.02	0.00
Echinoderm															-7.19	0.00	-5.52	0.00	-5.34	0.00	-1.09	0.28	-4.27	0.00	-3.44	0.38	-1.75	0.08	-5.52	0.00
Protochordata																	-6.10	0.00	-3.78	0.00	-0.12	0.91	-4.62	0.00	0.72	0.49	-0.75	0.46	-4.80	0.00
Arthropods																			-4.73	0.01	-0.05	0.51	-5.10	0.03	-2.43	0.28	-1.67	0.13	-5.87	0.00
Cnidaria																					0.77	0.47	-2.41	0.45	0.74	0.44	0.23	0.02	-3.27	0.01
Nematoda																							-0.32	0.68	-1.29	0.34	-0.35	0.73	-0.42	0.44
Early Eukarya																									-0.58	0.26	1.32	0.19	-2.83	0.02
Fungi																											2.49	0.36	-1.03	0.31
Bacteria																													-0.11	0.43

It is unclear whether fungi and early diverging metazoans possessed a one-protein (Nrf2 only) or two protein (Keap1-Nrf2) antioxidant response system. Nematodes having only a Nrf2-like sequence (329,436) are grouped with basal metazoans suggesting a one-protein system that likely evolved early in metazoan development. The position of nematode Nrf2-like sequences, however, differs from what would be expected from the generally accepted Tree of Life in which nematodes belong to a clade of "moulting animals" along with arthropods and several smaller phyla (437). The nematode Nrf2-like sequence SKN-1 also has a lower selective pressure (as measured by Codon-based Z-test of selection, Table 3.1) than the Nrf2 of most animals. Such, an alternative explanation is that recruitment of Keap1 had occurred shortly after Nrf2 evolution into an antioxidant response regulatory system at a time close to the animal-fungal divergence at late Stage 3 of atmospheric oxygenation. Homologous Keap1 proteins of nematodes may have subsequently lost persistence of regulatory control over Nrf2-like function in nematodes, perhaps due to a lack of environmental selective pressure attributed to the often hypoxic soil-dwelling lifestyle of worms. In contrast, tissues of cnidarians harbouring phototrophic endosymbionts can tolerate extremes of oxygen saturation (438), thus demanding efficient means to control oxidative damage. Accordingly, additional elaboration of Nrf2 activity in fungi and basal metazoans is essential to better elucidate evolutionary processes, which is enabled by the recent availability of several cnidarian genome annotations, including that of the scleractinian coral, *Acropora digitifera* (439).

In summary, we demonstrate that orthologues of Nrf2 first appeared in fungi around 1.5 Ga (408) during the Paleoproterozoic when photosynthetic oxygen was being absorbed into the oceans culminating in prolonged low oxidative stress (405). A subsequent significant divergence in Nrf2 is seen to occur during the split between fungi and the Metazoa approximately 1.0 – 1.2 Ga (413); at a time when oceanic ventilation released free oxygen to the atmosphere, but with most of this being absorbed by methane oxidation and oxidative weathering of land surfaces until approximately 800 Ma (422,428). Atmospheric oxygen levels thereafter accumulated during the Neoproterozoic giving rise to metazoan success during the Ediacaran period (635-541 Ma) leading to the Cambrian explosion (radiation) commencing at ~541 Ma (440). Atmospheric O_2 levels then rose in the late Paleozoic (359-252 Ma), driving further Nrf2 sequence divergence and Keap1 recruitment for Keap1-Nrf2 regulation of the oxidative stress response at the division between mammals and non-mammalian vertebrates during the during the Late Triassic (~225 Ma) (441,442). Understanding the evolution of Nrf2 and recruitment of other protein partners into an effective antioxidant response cascade might provide novel insights into the human ageing process since oxidative stress is believed to be

one of the key factors in ageing. This could, in turn, reveal possible new intervention strategies to improve metabolic health in our worldwide ageing population.

Chapter 4: Porphyra-334 and shinorine are antioxidants and antagonists of Keap1-Nrf2 binding

The content of this chapter has been submitted for publication as:

Gacesa R., Georgakopoulos N.D., Lawrence K.P., Yabe K., Dunlap W.C., Barlow D.J., Wells, G., Young A.R. and Long P.F. (2017). **Porphyra-334 and shinorine are antioxidants and antagonists of Keap1-Nrf2 binding.** FEBS letters (under review)

R.G. performed the extraction, purification and analysis of MAAs, performed the DPPH assay, interpreted the results and drafted the manuscript. The ORAC assay, fluorescence polarisation and thermal shift assays were performed jointly by R.G. and G.N.D.

4.1 Foreword to Chapter 4

The Chapter 2 of this thesis presented the structure based virtual screening study to determine the potential for competitive inhibition of Keap1-Nrf2 binding by natural products. The results presented in Chapter 2 demonstrated that the mycosporine-like amino acids (MAAs) have the capacity to bind to human Keap1-Nrf2 binding pocket. This chapter presents the study to identify if the MAAs are competitive inhibitors of human Keap1-Nrf2 interaction *in vitro.*

4.2 Abstract

Mycosporine-like amino acids (MAAs) are UV-absorbing metabolites typically produced by cyanobacteria and marine algae, but their properties are not limited to direct sun screening protection. Herein, we demonstrate that the MAAs porphyra-334 and shinorine are prospective activators of the cytoprotective Keap1-Nrf2 pathway as determined using fluorescence polarization and thermal shift assays to detect Keap1 receptor antagonism. Their in-vitro antioxidant activities determined by the DPPH free-radical quenching assay were low in comparison to ascorbic acid. However, their antioxidant capacity determined by the ORAC assay to quench free radicals via hydrogen atom transfer is substantial. Accordingly, the dual nature of porphyra-334 and shinorine to provide antioxidant protection offers a unique chemoprotective strategy to retard the progression of multiple degenerative disorders of ageing.

4.3 Introduction

The Kelch-like ECH-associated protein 1 (Keap1) is an actin bound homodimer that functions as a primary sensor of intracellular reduction-oxidation (redox) state regulation by controlling the activity of the master transcription nuclear factor erythroid 2–related factor 2 protein (Nrf2), which regulates the transcription of a large number of genes under control of the cis-acting enhancer termed the antioxidant response element (ARE) (157,443). The human Keap1 monomer is a 69.7 kDa protein composed of 625 amino acids that is divided into 5 distinct domains (Figure 1). The Kelch-repeat domain consists of six repeating motifs (KR1–KR6) that form a six-bladed β-propeller structure at which Keap1 binds to the Neh2 domain of Nrf2 (444). Under basal conditions, Nrf2 is targeted for ubiquitination and rapid 26S proteasomal degradation by Keap1 BTB domain bound Cullin3-Rbx1 E3 ubiquitin ligase (CRL^{Keap1}) (135,137,156). This turnover of Nrf2 prevents unnecessary expression of genes under Nrf2 transcriptional regulation (156). During conditions of oxidative stress, the ubiquitination and degradation of Nrf2 by CRL^{Keap1} is disrupted. Two separate models have been proposed for this dissociation: the "conformation cycling model" (137) and the "hinge and latch model" (135). In the conformation cycling model, it is proposed that, in the presence of cellular oxidants and exogenous electrophiles, covalent modification of Cys^{151} in the BTB domain of Keap1 causes conformational changes that prevent ubiquitination of Nrf2 by the CRL^{Keap1} protein complex (136,137). In the hinge and latch model, the Nrf2-Keap1 interaction is mediated by a high-affinity ETGE motif in the Neh2 domain of Nrf2, which functions as a "hinge" by stabilising Nrf2 binding to the Kelch domain in the Keap1 dimer. A low-affinity DLG motif in the Neh2 domain of Nrf2 functions as the "latch" by locking or unlocking the binding position of Nrf2, depending on the redox state of the cell. Under basal conditions, the DLG motif locks the Neh2 domain in the correct position to enable ubiquitination of Nrf2 (112,445). However, the IVR domain of Keap1 is cysteine rich and these residues are sensitive also to oxidation (134,446,447). During conditions of oxidative stress, these cysteine residues become oxidized to unlock the Nrf2 "latch", disrupting Nrf2 ubiquitination by the CRL^{Keap1} complex. As a consequence, Nrf2 is spared degradation at the proteosome, and newly translated Nrf2 proteins accumulate in the cell. Free Nrf2 then translocates to the nucleus where it forms a complex with small Maf proteins and interacts with the promoter region of the ARE to initiate the transcription of genes encoding a vast arsenal of proteins that protect against toxic contamination and regulate metabolic redox homeostasis (135,156,446).

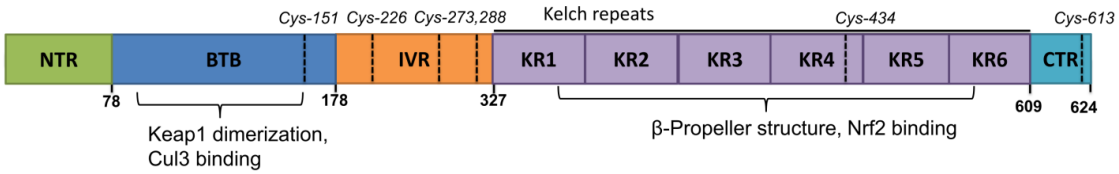


Figure 4.1 Illustration of the Kelch-like ECH-associated protein 1 (Keap1)

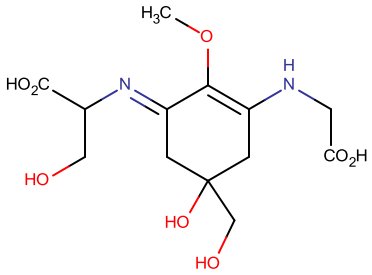
The Broad-complex, Tramtrack and Bric-a-Brac (BTB) domain, coloured blue, is responsible for the formation of the Keap1 dimer and for Nrf2 binding; the Kelch-repeat (KR) domain, forms a six-bladed β-propeller structure with DLG and ETGE motifs that bind with the Neh2 domain of Nrf2 (135,446,448). The Intervening (IVR) domain is comprised of amino acid residues between BTB and Kelch repeats. Cysteine residues that function as electrophile sensors are denoted in the above illustration.

Since oxidative stress has been implicated in numerous human diseases, the Keap1–Nrf2 protein-protein interaction (PPI) has become an important target for the potential development of therapeutic and chemopreventive agents. Numerous compounds have been examined for their ability to induce Nrf2-dependent gene expression, including those of natural origin (e.g., curcumin, sulforaphane) and others that are synthetic (e.g., bardoxolone methyl, oltipraz). Most of these Nrf2 activators are electrophiles that covalently modify the sulfhydryl groups of keap1 cysteine residues disrupting the ubiquitination and subsequent degradation of Nrf2 (448). These electrophilic inhibitors lack selectivity and thus increase the risk of "off-target" toxic effects due to indiscriminate reactions with cysteine residues in other cellular proteins. Accordingly, the discovery of direct, non-reactive, small molecule inhibitors of the Keap1–Nrf2 PPI appears to be the most promising strategy for Nrf2 activation to decrease the possibility of "off-target" toxic effects (449,450).

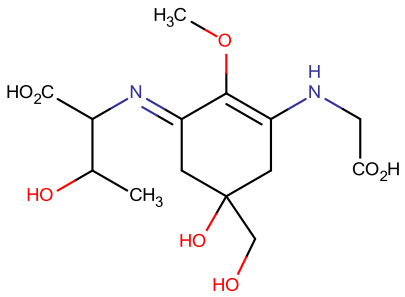
Mycosporine-like amino acids (MAAs) are small secondary metabolites commonly produced by marine algae and seaweeds that reside in shallow-water environments and are typically exposed to high levels of solar radiation. MAAs are found also in the tissues of some marine vertebrates, such as fish, that occur by dietary accumulation from the marine food-chain. MAAs absorb ultraviolet light, typically between 310 and 340 nm, allowing MAAs to protect cells from damaging solar UV radiation. Yet, MAAs are multifunctional metabolites that protect also against free-radical damage and boost cellular tolerance to desiccation, hyper-salinity and heat stress (64,451); there are more than 20 known MAAs in this class of natural metabolites. Previously, utilising protein modelling and virtual screening methods, we had predicted the potential for Nrf2 activation by competitive inhibition of its binding to Keap1, specifically by certain UV-protective MAAs (408). Here we provide *in vitro* empirical evidence to confirm our *in-silico* predictions that porphyra-334 (the principal MAA of *Porphyra*

yezoensis – Japanese seaweed "nori") and shinorine (from *Gloiopeltis furcata* – Japanese seaweed "fukuro-funori") may exert a cytoprotective function by specific, non-reactive binding to the Kelch-repeat domain of Keap1; additionally these MAAs are shown to have intrinsic antioxidant activity by quenching free oxygen radicals through hydrogen atom transfer.

A) Shinorine



B) Porphyra-334



Formula:	$C_{13}H_{20}N_2O_8$	$C_{14}H_{22}N_2O_8$
Mass:	332 Da	346 Da
λ_{max}:	332 nm	334 nm

Figure 4.2 Structures and biophysical characteristics of MAAs tested in this study.

Figure shows chemical structures of the mycosporine-like amino acids, shinorine (A) and porphyra-334 (B). The molecular formula, molecular mass, and the maximum wavelength of absorbance (λ_{max}) are given for each MAA.

4.4 Materials and Methods

4.4.1 Materials

All chemicals were purchased from commercial suppliers and used without further purification. Ascorbic acid, tBHQ, caffeic acid and DMSO were purchased from Sigma-Aldrich (Gillingham, Dorset, UK). Curcumin, DPPH, quercetin, and EGCG were purchased from Insight Biotechnology Ltd (Wembley, Middlesex, UK). Chlorogenic acid, trans-resveratrol and sulforaphane were purchased from Cambridge Bioscience Ltd (Cambridge, UK). Purified MAA compounds porphyra-334 and shinorine were kind gifts from Prof Kazuo Yabe.

4.4.2 Analysis of mycosporine-like amino acids

Identity and purity of mycosporine-like amino acid (MAA) samples was confirmed by HPLC analysis based on Carreto method (452), followed by the high resolution HPLC-MS/MS analysis to determine the mass and ion-fragmentation patterns of MAAs.

Each MAA sample was prepared by dissolving 1 mg of dry powder (previously stored at -20 °C in desiccator) in 1.5 mL of HPLC grade water. The solution was filtrated by 100-kDa ultrafilter (Ultra spin, Alltech) to remove water-insoluble materials and large molecules. HPLC analysis was performed on system composed of Shimadzu DGU-20A3 degasser, 2 x LC-20AD pump, CBM-20A communicator, SPD-M20A diode array detector and SIL-20A HT autosampler, using TPNA5011 Phenomnex C18 reverse phase column (Luna 5a 18C (2) 100A; 250 mm x 3 mm, 5 µm, No. 568381-1). 50 µL of sample was injected and elution was performed at 0.2 mL/min flow rate, using two solvent gradient elution, with Solvent A composed of HPLC grade water + 0.2% formic acid, pH corrected to 3.15 using NH_4OH, and Solvent B composed of 80% [HPLC grade water + 0.2% formic acid, pH corrected to 2.2 using NH_4OH and formic acid] : 10% Acetonitrile : 10% MeOH.

Table 4.1 Elution protocol for isolation of MAAs

The table lists the elution program for MAA isolation. Solvents A and B are described in the text.

Time (minutes)	Concentration (A)	Concentration (B)	Notes
0 – 2	100%	0%	
2 – 30	100% to 50%	0% to 50%	Linear gradient
30 – 35	50%	50%	
35 – 45	50% to 2%	50% to 98%	Linear gradient
45 – 50	2%	98%	
50 – 52	2% to 100%	98% to 0%	Linear gradient
52 – 60	100%		

High resolution mass spectrometry analysis was performed using HPLC system composed of Shimadzu DGU-20A3 degasser, 2 x LC-20AD pump, CBM-20A communicator, CTO-20A column oven, SPD-M20A diode array detector and SIL-20A HT autosampler coupled to the Bruker micrOTOF – qII mass spectrometer. The 50 µL of sample were injected into the system and elution was performed using the same solvents and program as the previous HPLC analysis (listed above).

4.4.3 Fluorescence polarization (FP) assay

The FP assay was carried out as previously described (159). Briefly, a solution of fluorescent peptide (FITC-β-DEETGEF-OH, 1 nM) and Keap1 Kelch domain (200 nM) in DPBS, pH 7.4 were mixed in an untreated black 96 well plate (Corning) with varying concentrations of test compounds up to 100 µM (final DMSO concentration 11 %, final volume 100 µL) and incubated for 1 hr at room temperature in the dark. FP was measured using a PerkinElmer

EnVision™ Multilabel Plate Reader. All measurements were recorded in triplicate. The normalised data were fitted to a standard dose-response equation by non-linear regression using Origin Pro software (OriginLab) to determine IC_{50} values.

4.4.4 Thermal shift assay

The thermal shift assay was carried out as previously described (294). Briefly, a solution of the detection dye SYPRO® orange (5X) and Keap1 Kelch domain protein (5 μM) in DPBS, pH 7.4 were mixed in a MicroAmp® Optical 96-well reaction plate (ThermoFisher) with varying concentrations of test compounds up to 100 μM (final DMSO concentration 10 %, final volume 40 μL). The plate was sealed using an optical adhesive cover and wrapped with aluminium foil to protect the dye from light. The plate was then transferred to a plate centrifuge and spun briefly (200 x *g*, 1 min, at room temperature) to remove any air bubbles and to collect the reaction mixture at the bottom of the wells. The plate was incubated for 1 hr at room temperature and then placed into a 7500 Real Time PCR machine and heated using a standard protein melting protocol (453). The fluorescence intensity was recorded with excitation at 465 nm and emission measured at 580 nm during a temperature scan from 25 °C to 95 °C with a temperature ramping rate of 1 °C/min. All measurements were performed in triplicate. The raw data were exported to MS Excel and analysis was performed using a custom script provided by Structural Genomics Consortium, University of Oxford. The temperature range over which protein unfolding occurred was established at temperatures below the maximum fluorescence intensity. The processed data were fitted to the Boltzmann equation by linear regression analysis using Origin Pro software.

4.4.5 DPPH radical scavenging activity

The percentage of antioxidant scavenging activity for each MAA was determined according to methodology described by Mishra *et al.* (454). Briefly, each MAA (62.5 μM, 117.7 μM and 210.5 μM concentrations) and an ascorbic acid positive control (3.9 μM, 7.8 μM, 15.6 μM, 31.3 μM, 62.5 μM and 312.5 μM concentrations) were prepared in DMSO. For each sample, the reaction mixture consisted of 0.1 mL of the test sample and 1.5 mL of 70 μM DPPH in methanol. The colour change from violet to yellow, when DPPH is reduced upon reaction with an antioxidant, was recorded at 515 nm using a UV/VIS spectrophotometer (model 7315, Jenway Ltd., Stone, Staffordshire, UK) after 30 mins of reaction at room temperature, with reaction mixtures shielded from light. The mixture of DMSO (0.1 mL) and DPPH (1.5 mL) served as the reaction blank. The percentage of DPPH radical scavenging activity was calculated as: % *scavenging activity* $= 100 \times (A_{blank} - A_{sample})/A_{blank}$. The DPPH scavenging

activity was found to be proportional to the MAA concentration (for levels examined), and the 515 nm absorbance of the fully reduced DPPH was set to zero. Experiments were performed in technical triplicates with three replicates, and % scavenging activities were plotted as the mean of the 9 triplicate/replicate values against test sample concentrations (μM). IC_{50} values were estimated from linear regression of the data.

4.4.6 ORAC antioxidant assay

The oxygen radical absorbance capacity (ORAC) for each MAA was carried out using the ORAC Antioxidant Assay Kit (Zenbio, North Carolina, USA) according to manufacturer's instructions. Trolox standards were prepared in the assay buffer (0 - 100 μM) along with serial dilutions of each MAA (500 - 0 μM) and an ascorbic acid positive control (0 - 100 μM). 150 μl of the fluorescein working solution was added to central wells of a 96 well plate, with 25 μl of each of the standards or MAA in duplicate, and the plate incubated at 37 °C for at least 15 mins. The 2,2'-azobis(2-amidinopropane) dihydrochloride (AAPH; sold as 2,2'-azobis(2-methylpropanimidamine) dihydrochloride) working solution was added to each well (25 μl) to start the reaction. Fluorescence was measured in a preheated incubation chamber (37 °C) using a Spectra Max 384 Plus spectrophotometer with excitation/emission = 485/530 nm immediately (t = 0) and then every minute for 30 mins. Standard curves were generated for each compound and the area under the curve calculated. Each MAA tested was expressed as a Trolox equivalent concentration.

4.5 Results

4.5.1 HPLC and MS analysis of MAA samples

The purity and identity of mycosporine-like amino acids shinorine and porphyra-334 (Figure 4.2) was confirmed based on the HPLC retention times, comparison with Helioguard-365 which is known to contain shinorine, palythine and porphyra-334 (455), and by high resolution tandem mass spectrometry. The results are presented in Figure 4.3 and Figure 4.4.

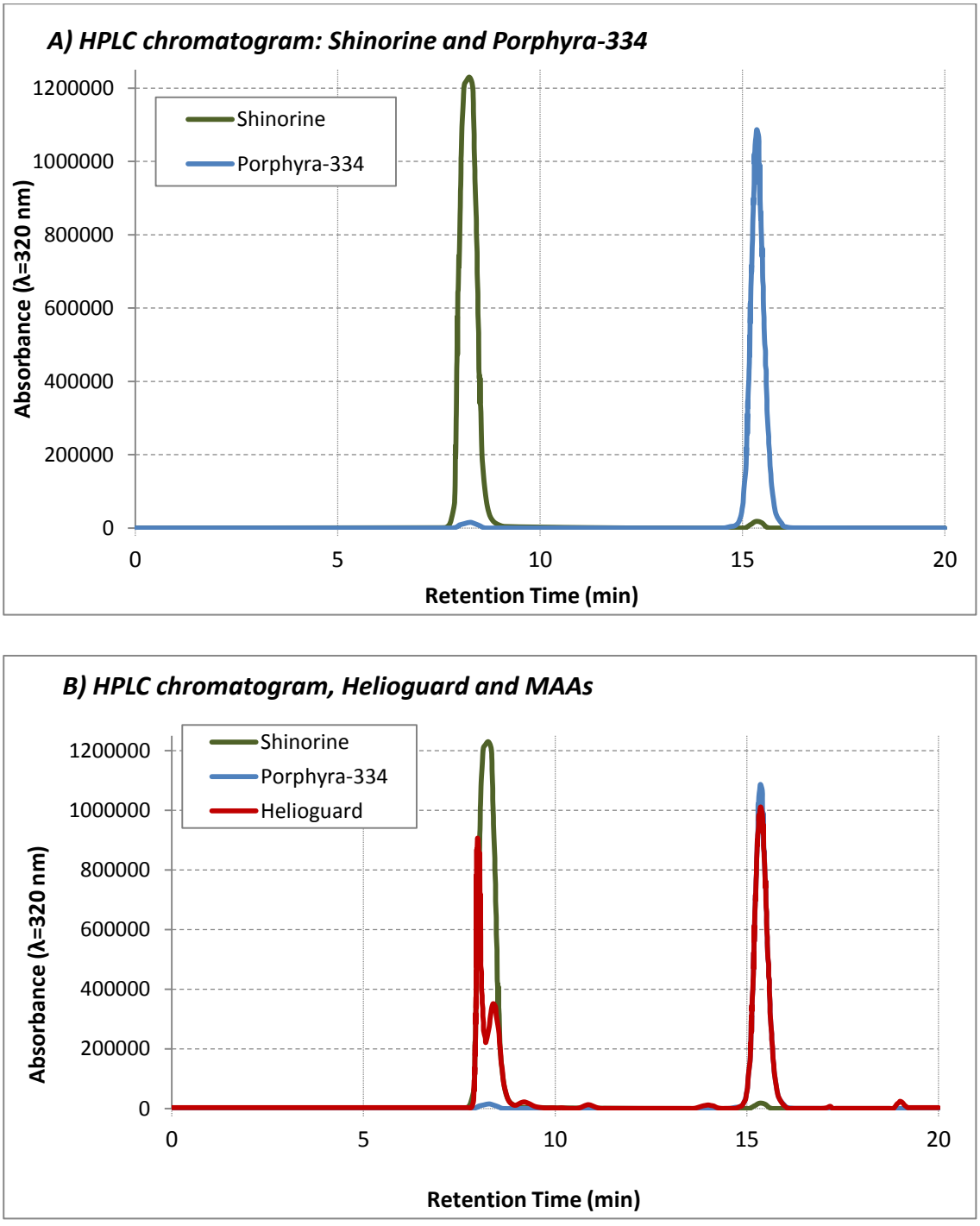


Figure 4.3 HPLC chromatograms of MAA samples and Helioguard

Figure A) displays HPLC chromatogram of MAA samples Shinorine (green) and Porphyra-334 (blue), while figure B) shows chromatogram of MAA samples mixed with Helioguard 365 (red), commercially available extract of red alga *Porphyra umbilicalis* known to contain Porphyra, Shinorine and Palythine.

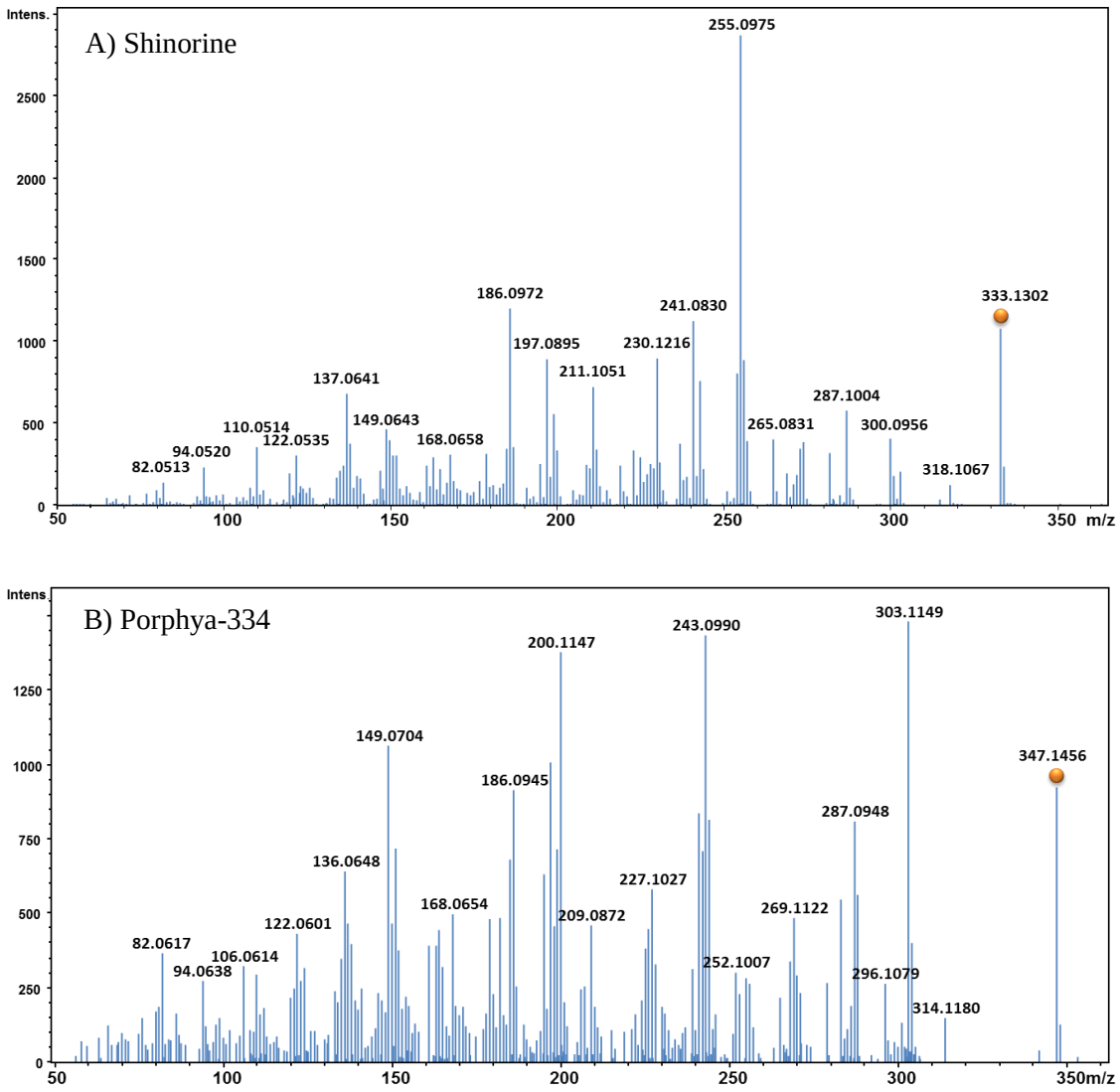


Figure 4.4 MS/MS fragmentation patterns of shinorine and porphyra-334

The figure A) displays the fragmentation patterns of ionized shinorine, while figure B) shows the MS/MS spectrum of porphyra-334. Masses of major MS/MS fragments are labelled and masses of the precursor ions are marked with orange circles (m/z = 333.1302 for shinorine and 347.1456 for porphyra-334).

4.5.2 Keap1-binding activity of MAAs

The MAAs porphyra-334 and shinorine together with eight selected antioxidants (ascorbic acid, tBHQ, caffeic acid, chlorogenic acid, curcumin, quercetin, EGCG, trans-resveratrol) plus the known electrophilic Nrf2 activator sulforaphane were evaluated for their interaction with the Kelch-repeat domain of the Keap1 protein using both *in-vitro* FP and thermal shift assays. In the FP assay, the test compounds competed with a fluorescein labelled peptide **(**fluorescein-[β-ala]DEETGEF-OH) based on the high affinity ETGE motif that binds Nrf2 to the Kelch-repeat domain (Figure 4.5). The protein-protein interaction of the fluorescein-[β-Ala]-DEETGEF-OH

peptide had an IC_{50} of ~3 μM in the presence of the Kelch-repeat domain protein (159). Both porphyra-334 and shinorine gave nearly equivalent ligand-receptor IC_{50} values of ~100 μM, requiring a much greater concentration in the presence of excess Kelch-repeat domain protein than the binding of the high affinity, labelled synthetic peptide (159). In contrast, no significant interactions were detected between the eight known antioxidants or sulforaphane and the Kelch-repeat domain protein.

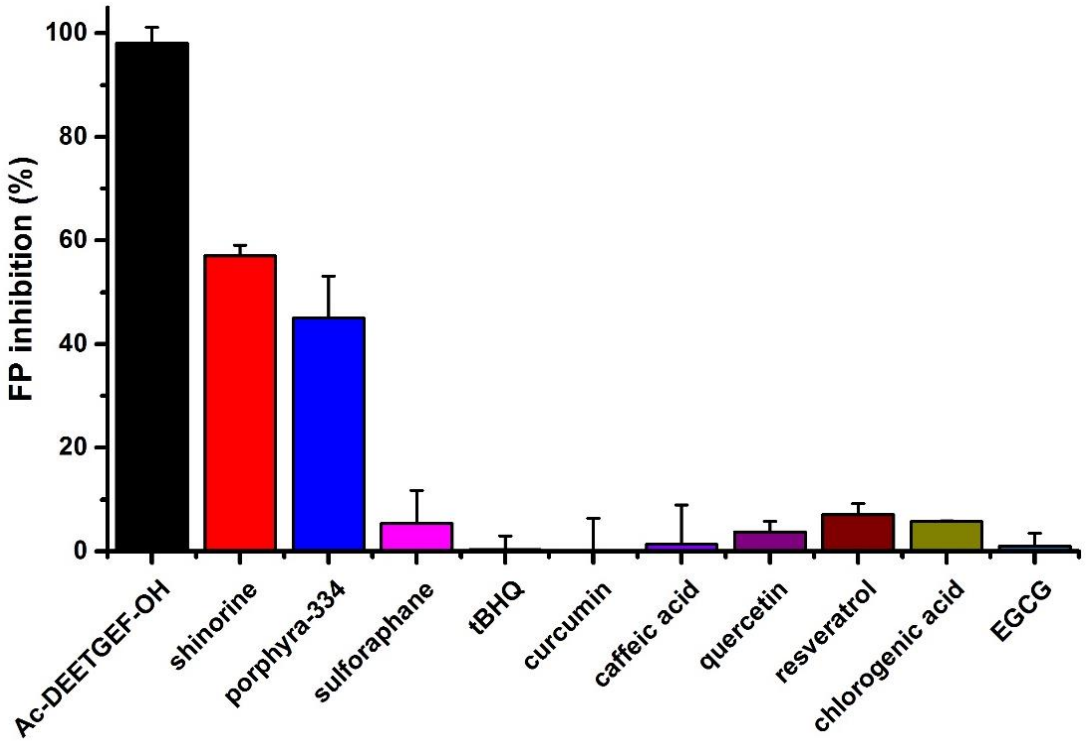


Figure 4.5 Fluorescence polarization measurement of specific, non-reactive binding of MAAs to the Kelch-repeat domain of Keap1.

Ligand-receptor binding was measured using a fluorescence polarization assay in comparison to a high affinity, labelled synthetic peptide. The FP inhibition (in percent) is displayed for small peptide positive control, MAAs shinorine and porphyra-334 and range of electrophile Nrf2 activators, at the compound concentration of 100 μM.

The MAAs and eight selected antioxidants were then tested as ligands of the Kelch-repeat domain protein using the thermal shift assay. The assays involved treatment of the Kelch-repeat domain protein with these test compounds of interest, followed by denaturation by heating to separate the protein-ligand aggregates from the soluble protein fraction. Whereas unbound proteins denature and precipitate at elevated temperatures, ligand-bound proteins remain in solution. An unlabelled version of the peptide ([β-ala]DEETGEF-OH) used in the FP assay was used as the positive control in the thermal shift assays against which binding of the test compounds to the Kelch-repeat domain were compared. When used at 50 μM concentration,

the [β-Ala]-DEETGEF-OH peptide had a high ΔTm of 3.91 ± 0.04 °C. Both porphyra-334 and shinorine also demonstrated elevated binding values to the Kelch-repeat domain protein, albeit when used at 100 μM (porphyra-334 ΔTm = 0.93 ± 0.02 °C, shinorine ΔTm = 0.64 ± 0.11 °C). However**,** no significant protein-ligand interactions were detectable when the 8 antioxidants or the electrophilic Nrf2 activator sulforaphane were tested. The thermal shift assays reflected the results of the FP assays for binding of MAAs and test substrates to the Kelch-repeat domain protein of Keap1.

4.5.3 Antioxidant activity of MAAs

Scavenging of the DPPH free radical is a common assay used to assess the antioxidant activity of a compound. Ascorbic acid is a well-known DPPH free-radical scavenger and is commonly used as a positive control, which was used also for comparison in this study. The IC_{50} value for ascorbic acid was 24.5 ± 1.1 μM in the methanol reaction medium. The free radical scavenging activities of MAAs were markedly lower compared to that of ascorbic acid. The IC_{50} values were: 185.2 ± 3.2 μM for porphyra-334 (equivalent to 13.23 % of the activity of ascorbic acid) and 399.0 ± 1.1 μM for shinorine (equivalent to 6.14 % of the activity of ascorbic acid).

Another method commonly used to assess antioxidant activity is the ORAC assay, which is used to assess the hydrogen atom transfer capacity of a test compound to suppress peroxyl radical (ROO•) induced oxidative damage measured by the fluorescence decay of fluorescein. Trolox, a water-soluble vitamin E analogue, is used as the standard against which other compounds are compared. Ascorbic acid was used also as an additional control reference. Both MAAs demonstrated significant activity (Figure 4.6) with porphyra-334 displaying an antioxidant capacity equal to 39.4 ± 9.4 % of equimolar Trolox and shinorine showing 12.2 ± 6.12 % of the Trolox capacity. Ascorbic acid showed greater antioxidant capacity than did Trolox (equivalent to 126.9 ± 13.1 % of Trolox) and that of our test MAAs.

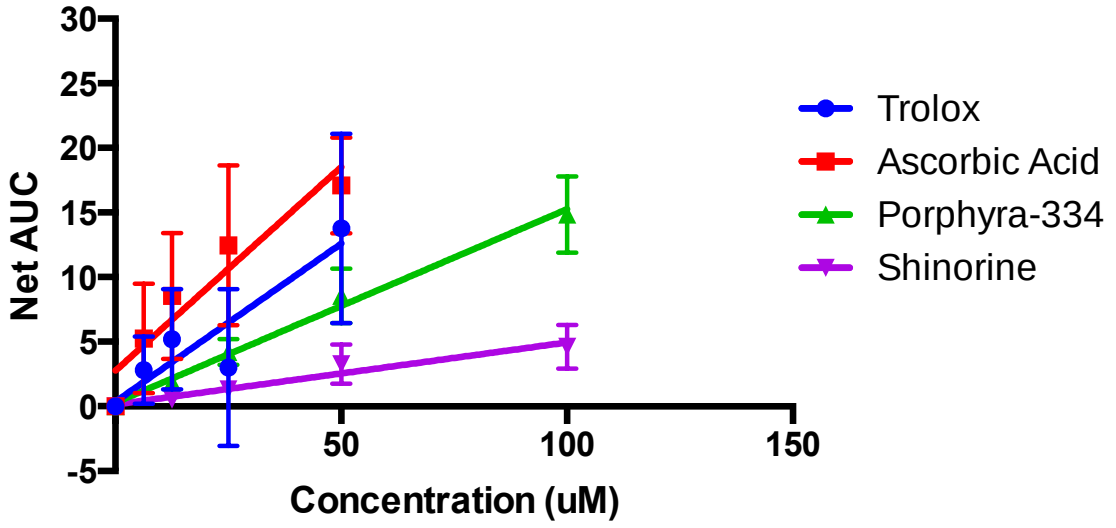


Figure 4.6 The oxygen radical absorbance capacity of MAAs.

MAAs and reference controls of increasing concentration were assessed for their ability to quench the ROO• radical as a measure of antioxidant activity. All compounds demonstrated significant activity (trolox p=0.0031; ascorbic acid p=0.0133; porphyra-334 p=0.0002; shinorine p=0.0044; all R^2>0.9000 determined by linear regression analysis). Relative antioxidant capacity compared to trolox was calculated (ascorbic acid = 126.9 ± 13.1 %; porphyra-334 = 39.4 ± 9.4 %; shinorine = 12.2 ± 6.12 %). Each data point represents the mean value (n=3).

4.6 Discussion

Aerobic organisms are regularly challenged by various environmental oxidants and toxic electrophiles that include free radicals produced from exposure to UV, visible and IR radiation, xenobiotic pollutants and endogenously generated, reactive by-products of oxidative metabolism. Such, these organisms have long evolved a robust detoxification and redox signalling systems to prevent cellular damage (112,156). The transcription factor Nrf2, targeting ARE upstream promoter sites of many protective genes, is the principal regulator of mammalian cellular antioxidant defence and P-450 cytochrome detoxification proteins. While various intrinsic pathways and signalling cascades have been identified that contribute to the regulation of Nrf2 function (156,456), the key role of the repressor Keap1 protein is firmly established since disruption of Keap1-Nrf2 binding is sufficient to increase significantly nuclear Nrf2 activity (151,457). This is additionally supported by siRNA knockdown of Keap1 mRNA (152) and Keap1-null ablation (151), both of which massively enhance free cellular Nrf2 accumulation. Furthermore, somatic mutations of Keap1, a characteristic of certain cancer phenotypes (153,217), result in constitutive Nrf2 activation, as does the silencing of Keap1 expression by miRNAs (175,458), or by epigenetic hyper-methylation of the Keap1 promotor (155).

Human Keap1 contains six ROS/electrophile-sensing cysteine residues (Cys151, Cys226, Cys273, Cys288, Cys434 and Cys613) which are prone to oxidation and covalent modification by electrophiles. These changes do not prevent Keap1-Nrf2 binding, but alter the basal-state "open" conformation, which prevents CRL^{Keap1} ubiquitination to release Keap1 upon Nrf2 degradation (136). As a result, free Keap1 is not regenerated, and newly synthesized Nrf2 accumulates to promote gene transcription. Electrophile activators of Nrf2 and the majority of other known Nrf2 inducers, including sulphoraphane used in this study, function also by this mechanism of targeting critical Cys residues of Keap1 to activate Nrf2 transcription (156,459). The compounds examined in this study, with the exception of the Keap1-binding MAAs, are sometimes referred to as "indirect antioxidants" or "Nrf2-activating antioxidants" and would not be expected to show appreciable activity in the FP, thermal shift, DPPH or ORAC assays. These expectations were matched by our assay results.

Non-electrophile activators of Nrf2, e.g. small-peptides (158,159) and small-molecule Nrf2 activators (160,460), function by competitive inhibition of Keap1-Nrf2 binding. These compounds hinder the formation of the "closed" Keap1-Nrf2 conformation by binding to the Keap1 β-propeller structure responsible for DLG and ETGE binding. While less studied than electrophile-based Nrf2 activation, targeting of the Keap1-Nrf2 binding site by competitive inhibitors has potential to upregulate Nrf2 activity without the toxic danger of indiscriminate "off-target" effects. These compounds and peptides do not depend on covalent electrophile modification of protein cysteine residues and are prone to indiscriminate reactions with proteins of other cellular signalling networks (157). In this study we present empirical *in vitro* evidence for the MAAs porphyra-334 and shinorine to competitively interact with the Nrf2 binding site of the human Keap1 protein, as previously anticipated from bioinformatic, structure-based Kelch domain binding predictions (408).

The first study to examine the antioxidant properties of MAAs, using the phosphatidylcholine peroxidation inhibition assay, found that imino-MAAs (shinorine, porphyra-334, palythine, asterina-330 and palythinol) were oxidatively robust to AAPH-generated peroxyl radicals, whereas the oxo-MAA mycosporine-glycine strongly inhibited radical-initiated phosphatidylcholine autoxidation in a concentration-dependent manner (250). 4-Deoxygadusol, presumed to be the immediate biochemical precursor of MAAs, was found also to have strong antioxidant properties (244) as demonstrated by voltamperogramic comparison of the electrochemical properties of 4-deoxygadusol, mycosporine-glycine and mycosporine-taurine (64). Numerous studies on the antioxidant properties of MAAs followed thereafter and have been extensively reviewed (232). Notably, the antioxidant capacities of porphyra-334 and

shinorine, together with other MAAs isolated from marine extracts (three seaweeds and one lichen), were compared using the ABTS radical-cation decolorization and superoxide-scavenging pyrogallol autoxidation assays, which showed scarce activity in comparison to mycosporine-glycine. Yet, shinorine and porphyra-334 showed modest antioxidant activity using the β-carotene/linoleate bleaching assay (251). Contrasting reports by Rastogi et al. attribute significant antioxidant activities of cyanobacterial extracts containing a mixture of imino-MAAs, albeit the IC_{50} values of these MAAs were approximately 15 – 30 % that of the antioxidant activity of ascorbic acid as measured by DPPH-radical, ferric-reducing antioxidant power (FRAP) and superoxide radical scavenging activity (SRSA) assays (253,254). We report also that purified shinorine and porphyra-334 showed low antioxidant activity in the DPPH radical assay (252) but exhibited significant antioxidant activity using the ORAC assay, which is a measure of the oxygen radical absorbance capacity of a test substance.. Previously, only the MAA precursor gadusol had been examined using this assay, and its antioxidant capacity was found comparable to that of ascorbic acid (249). Our DPPH assay findings are consistent with established data (232), whereas our ORAC assay results reveals that both pophyra-334 and shinorine have significant antioxidant capacity by quenching oxygen radicals by hydrogen atom transfer, rather than by reductive electron transfer (461).

Several recent publications describe the immunoregulatory and anti-inflammatory properties of mycosporine-glycine, shinorine and porphyra-334 (252,462,463). Suh et al. (252) reported that treating the human fibroblast cell line HaCaT with mycosporine-glycine resulted in a significant decrease in UV-induced, inflammatory COX-2 mRNA levels, whereas all MAA treatments increased expression levels of the UV-suppressed, procollagen C proteinase enhancer and elastin genes. Becker et al. (462) found that both shinorine and porphyra-334 induced nuclear factor NF-κB activity in the NF-κB/AP-1 reporter myelomonocyte cell line THP-1-blue, although NF-κB induction was greater with shinorine. Yet, while shinorine marginally enhanced LPS-superinduced NF-κB activity, porphyra-334 significantly reduced the NF-κB response in LPS-stimulated cells. Ryu et al. (463) had reported previously that porphyra-334, in concentration as low as 10μM, had an inhibitory effect on the expression of NF-κB dependent inflammatory genes, such as IL-6 and TNF-κ, in UVA-irradiated skin fibroblasts, but pertinent to this study is finding that porphyra-334 can activate the Nrf2 signalling pathway in UVA-irradiated cells. However, nuclear Nrf2 translocation by porphyra-334 without concurrent ROS production by UVA exposure had not been demonstrated, and it is thus unknown whether the Nrf2 activating activity of porphyra-334 is a result of direct inhibition of Keap1-Nrf2 binding or if porphyra-334 is an electrophile activator of Nrf2 and

activates cytoprotection by modifying critical cysterine residues of Keap1. Results of our ORAC and DPPH assays (Figure 4.6, Figure 4.5) and previously published studies (232) suggest that porphyra-334 is an antioxidant, and is thus unlikely to have pro-oxidant activity required to oxidise the critical Cys residues of Keap1. The products of MAA metabolism in human cells is are, however, currently unknown and it is possible that porphyra-334 is metabolised into electrophile activators of Nrf2 *in vivo*. In addition, while our previous virtual screening experiments and *in-vitro* results presented in this study (Figure 4.5) demonstrate that porphyra-334 is a direct inhibitor of Keap1-Nrf2 interaction *in vitro*, it is currently unknown if porphyra-334 also interacts with other Keap1-like proteins such as β-TrCP protein, an indirect modulator of Nrf2 activity. Thus, metabolism of MAAs and interaction of MAAs and other β-propeller proteins are avenues worthy of future research.

In conclusion, results presented herein lend support to our findings that porphyra-334 and shinorine are prospective activators of the cytoprotective Keap1-Nrf2 pathway by Keap1 receptor antagonism.

Chapter 5: Proteomics approach to study the UV-induced oxidative stress in yeast isolated from environments with high incidence of solar radiation

5.1 Foreword to Chapter 5

The phylogenetic studies of Keap1-Nrf2 pathway, presented in chapters 2 and 3, have identified sequences encoding the proteins similar to the vertebrate transcription factor Nrf2 in genomes of all major animal phyla and certain fungi. The study presented in this chapter used the multidimensional protein identification technology to quantify the protein expression level fold changes in the UV-tolerant yeast model exposed to long term UV-B irradiation. The yeast proteome was examined to identify if the yeast model expresses bZip proteins homologous to the animal Nrf2, as was predicted in studies presented in chapters 2 and 3. In addition, the quantified yeast proteome was annotated using biological databases to describe the UV-stress response of the UV-tolerant yeast isolated from environment exposed to high incidence of solar radiation.

5.2 Introduction

Ultra violet radiation (UVR) is a major hazard to biological systems and an important environmental source of oxidative stress. The UV-B band of UVR (λ = 280-315 nm) inflicts direct damage to cellular macromolecules by photo-oxidation of proteins (38) and by inducing the formation of DNA photo-adducts (37). In addition, both UV-A (λ = 315 – 400 nm) and UV-B radiation cause oxidative stress by inducing the production of reactive oxygen species (RS) in irradiated cells (39,40). The basic leucine zipper domain (bZip) transcription factor Nrf2 and its inhibitor, the kelch-domain protein Keap1, play critical roles in regulation of response to oxidative stress in vertebrates (464). In an unstressed cell, Nrf2 is sequestered in the cytosol by Keap1, ubiqutinated by the Cullin3-Rbx1 E3 ubiquitin ligase – Keap1 (CRL^{Keap1}) complex and proteosomally degraded. During oxidative stress, the increased cellular concentration of RS causes the oxidation of critical, "RS sensing", cysteine residues of Keap1, which leads to conformational changes in the CRL^{Keap1} complex and prevents ubiquitination of the Nrf2 protein (156). The inhibition of the CRL^{Keap1}-mediated degradation of Nrf2 results in an increase in cytosolic concentration of newly synthesized Nrf2 and its translocation to the nucleus (137). The nuclear Nrf2 associates with small Maf proteins, and interacts with the

antioxidant response element (ARE) cis-acting enhancer sequences on the DNA to activate the transcription of a large number of genes that encode detoxification and antioxidant proteins (465,466).

In addition to playing a major role in resistance to environmental pollutants (183) and chemically induced oxidative stress (186), the Nrf2 pathway also regulates response to UV-induced oxidative stress (467). In a study by Kleszczyński et al, the induction of Nrf2-regulated genes by pre-treatment with Nrf2 activators sulforaphane and phenylethyl isothiocyanate was shown to protect human skin cells against UV exposure in cell cultures and *ex vivo*. In this study, the pre-treated, UV-exposed, cells had lower rates of apoptosis, reduced levels of biomarkers of sunburn, and higher levels of endogenous antioxidants such as catalase, when compared to non-treated controls (468). The Nrf2-mediated protection against UV-induced oxidative stress was also shown in mouse models, where treatment with sulforaphane-containing broccoli sprouts significantly reduced the incidence rate of UV-induced skin cancer (469). Furthermore, a study with genetically modified mouse models found that mouse models lacking a Nrf2-encoding gene recover slowly from the UV-induced inflammation and are highly sensitive to photo-ageing (470).

In the model yeast *Saccharomyces cerevisiae*, UV irradiation activates the RAS/cAMP/PKA signalling pathway independent of DNA damage. This pathway includes the GTPase RAS which controls the activity of adenylate cyclase (Cyr1) enzyme which stimulates the production of cyclic AMP (cAMP). The increased intracellular concentration of cAMP triggers the activation of cAMP-dependant protein kinase A (PKA), and initiates a phosphorylation cascade that activates the bZip transcription factor Gcn4 (471). Gcn4 is primarily associated with yeast response to starvation and controls the transcription of more than 30 genes that encode enzymes involved in biosynthesis of amino acids (472). The study of genetically modified *S. cerevisiae* strains by Engelberg et al (1994), however, found that *S. cerevisiae* mutants with high, constitutive, expression of Gcn4 are ~3.5 fold more resistant to UV irradiation compared to wild type yeasts, while the strains deficient in Gcn4-encoding gene are ~5-fold more sensitive to UV than wild type yeasts (471). In addition to RAS signalling, the other major UV-response pathway in *S. cerevisiae* is the Yap1-mediated response to UV-induced oxidative stress (473). The yeast transcription factor Yap1 is a bZip protein that binds to the AP-1 recognition element in the promotor regions of numerous yeast genes involved in the DNA repair and response to oxidative stress, such as glutathione biosynthesis enzymes and thioredoxins (474). While Yap1 signalling is reminiscent of the Nrf2-mediated activation of response to oxidative stress in vertebrates, no interaction between kelch-like proteins and Yap1 has been reported as of yet

(475). In addition to RAS and Yap1 stress response signalling pathways, yeasts also possess DNA damage response pathways. These pathways are largely conserved across all eukaryotes and consist of sensors, primary and secondary signal transducers and effectors. The DNA damage is detected by "sensor" proteins, such as *S. cerevisiae* proteins Rad1, Rad9 and Hus1, and activates the signal transduction kinases (Mec1 and Tel1 in yeast *S. cerevisiae*) (476), which activate the secondary, downstream kinases such as Dun1 (477). The downstream kinases activate the currently unknown effector proteins that increase the transcription rates of enzymes involved in DNA repair and also arrest the cell cycle during the DNA repair (476).

While the baker yeast *S. cerevisiae* is highly resistant to UV radiation, when compared to bacterial models (478), yeasts native to the environments with high incidence of solar radiation can tolerate the UV radiation levels lethal to *S. cerevisiae*. Examples of such yeasts include carotenogenic yeasts of the *Sporobolomyces* and *Rhodotorula* Genera (478,479), and the black yeasts of Genus *Exophiala* (478). The mechanisms of response to UV-induced stress have not been studied in these, UV-tolerant, yeasts, and it is currently not known if the stress response signalling, mediated by bZip proteins, is conserved between UV-tolerant yeasts, baker yeast *S. cerevisiae* and the animal Keap1-Nrf2 pathway. The UV-tolerant, *Sporobolomyces*, yeast model was selected as a model organism in this study because of its taxonomical classification (Phylum Basidiomycota as opposed to *S. cerevisiae* which belongs to Phylum Ascomycota), and to allow a study of effects of high levels of UV-B in range lethal to the *S. cerevisiae*.

The aim of this study was to determine, at the proteome level, stress response mechanisms of the UV-tolerant yeast model isolated from the leaves of plants collected from the São Paulo region in Brazil, regularly exposed to high incidence of solar radiation (UV index ~13.9) (480). The yeast model, identified as *Sporobolomyces sp.* and designated LEV-2, was exposed to long-term UV-B irradiation, and proteomes of LEV-2 cultures exposed to different durations of UV-B (5 minutes to 24 hours) were quantified by MudPIT technology. The quantified proteins were functionally annotated using gene ontology (GO) terms (481), KEGG modules and KEGG pathways (342,482) to identify proteins involved in stress response and bZip transcription factors. In addition, the fold changes of proteins involved in stress response were examined to describe the stress response of the UV-tolerant yeast LEV-2.

5.3 Methodology

5.3.1 Isolation of UV-tolerant yeasts from environmental samples

Seven leaf samples were collected from the city campus of the University of São Paulo, from the following plants: Goaiba (*Psidium guajava*), Jacarandá (*Jacaranda cuspidifolia*), Ipê Amarelo (*Tecoma serratifolia*), Quaresmeira (*Tibouchina granulosa*), Mangueira (*Mangifera indica*), Tipuana (*Tipuana tipu*) and Jamelão (*Syzgium cumini*). The leaves were collected from the plants exposed to mid-day sun during summer (mid. January, predicted UV index ~13.9 (480)). Collected leaves were cut into 0.5 x 0.5 cm pieces, and each piece was placed onto the surface of a Petri dish containing half-strength yeast-peptone-dextrose (YPD) solid media with antibiotics, composed of yeast extract (5 g/L), dextrose (10 g/L), peptone (10 g/L), agar (6 g/L) and antibiotics (100 mg/L chlortetracycline, 20 mg/L chloramphenicol and 20 mg/L streptomycin-sulphate, added into the media after the sterilization).

The surfaces of Petri dishes were irradiated for 1 minute, using dual Philips Ultraviolet-B TL 20W/12RS lamps, with UVR output of UV-B: 4 $J/m^2/s$ and UV-A: 1.75 $J/m^2/s$. The UVR output of lamps was measured at 10 cm distance from the lamps, using a Bentham (double grating) DM150BC spectroradiometer with 2 400 g/mm grating blazed at 250 nm, IS4C integration sphere diffuser, DH3 (bi) photomultiplier and large aperture. Results of spectral analysis of UV lamps are presented in Appendix C-1. Samples were irradiated in open petri dishes, at a distance of 10 cm from the lamps.

Irradiated samples were incubated for 24 hours at 25 °C, and yeast colonies (if any) were transplanted into sterile petri dishes containing half-strength YPD media with antibiotics. The process was repeated daily for 7 days, and yeast colonies were transplanted as necessary to isolate pure strains. The strains were considered pure if a single, distinct, colony could be isolated. The isolated microorganisms identified as yeasts (based on morphology and resistance to antibiotics) were retained for further analysis.

5.3.2 UV-tolerance testing of yeast isolates

The isolated yeast samples, the baker yeast *Saccharomyces cerevisiae*, and five UV-tolerant yeasts previously isolated and identified by Casteliani et al. (479), designated LEV-2, LEV-9, LEV-12, LEV-13 and LEV-16, were tested for survival after exposure to long term UV-B irradiation. The LEV yeast isolates were kindly provided by Prof Itamar Soares (Laboratory of Environmental Microbiology, Embrapa Environment, Jaguariúna, SP, Brazil). The yeasts were cultivated in sterile half-strength YPD liquid medium, composed of yeast extract (5 g/L),

dextrose (10 g/L) and peptone (10 g/L), in cotton-plugged 250 mL Erlenmayer flasks, at 27 °C with shaking (100 rpm). Three replicates were grown for each sample. After 24 hours of growth, optical density at 600 nm (OD_{600}) of each sample was standardized to OD_{600} = 1.0 (~3 x 10^7 cells / mL), by diluting the sample with sterile half-strength YPD liquid medium as necessary, and 25 mL aliquot of each yeast culture was taken for UV-tolerance testing. The yeast samples were irradiated by dual Philips Ultraviolet-B TL 20W/12RS lamps as follows:

- Sample 1 (Control) was not exposed to UV
- Sample 2 (1h) was irradiated for 1 hour
- Sample 3 (2h) was irradiated for 2 hours
- Sample 4 (4h) was irradiated for 4 hours
- Sample 5 (8h) was irradiated for 8 hours
- Sample 6 (24h) was irradiated for 24 hours

Irradiation was performed at room temperature and temperature changes during the irradiation were assumed to be moderate and not critical for the outcome of experiment. During the irradiation, each sample was standardized to volume of 25 mL by adding sterile half-strength YPD media to account for any evaporation during the irradiation. The irradiated samples were vortex agitated for 30 seconds, and 1 mL aliquots were used to prepare three technical replicates of ten-fold serial dilutions. The 0.1 mL of the sample diluted 1:1000, containing ~3.0 x 10^3 cells, was used to inoculate half-strength YPD solid agar. Inoculated petri dishes were incubated at room temperature, and yeast colonies were counted after 48 hours. Survival curves were constructed from the mean values of the replicates, and the survival rates were calculated as SR (%) = 100 x [CFU (irradiated sample) / CFU (control sample)].

5.3.3 Preparation of UV-tolerant yeast isolate for proteomics

The yeast culture was cultivated in half-strength YPD media for 24 hours. After 24 hours, cell numbers were estimated at OD_{600} and samples diluted to OD_{600} = 1.0, equivalent to approximately 3 x 10^7 cells/mL. Measurements were performed using a UV/VIS spectrophotometer (model 7315, Jenway Ltd., Stone, Staffordshire, UK).

Yeast culture was then divided into 20 x 25 mL aliquots (Sample 1 – Sample 10, in duplicate). The samples, in open petri dishes, were irradiated under dual Philips Ultraviolet-B TL 20W/12RS lamps (UVR output of UV-B: 4 J/m^2/s and UV-A: 1.75 J/m^2/s) at a distance of 10 cm from the lamps. The samples were stirred manually every 20 minutes during irradiation.

Yeast samples were irradiated as follows:

- Sample 1 (Control) was not exposed to UV
- Sample 2 (5') was irradiated for 5 minutes
- Sample 3 (10') was irradiated for 10 minutes
- Sample 4 (15') was irradiated for 15 minutes
- Sample 5 (1h) was irradiated for 1 hour
- Sample 6 (2h) was irradiated for 2 hours
- Sample 7 (4h) was irradiated for 4 hours
- Sample 8 (8h) was irradiated for 8 hours
- Sample 9 (24h) was irradiated for 24 hours
- Sample 10 (Control 2) was not exposed to UV and was kept on the lab bench (artificial light, room temperature, no stirring) for 24 hours

During the irradiation, each sample was standardised to volume of 25 mL by adding sterile half-strength YPD medium to account for any evaporation. Samples were vortex agitated for 30 seconds, and three 1 mL aliquots were taken for DPPH antioxidant assay. The yeast cells in the remaining culture (22 mL) were collected by centrifugation at 1 000 x g for 15 minutes at room temperature. The supernatant was discarded and the cells transferred to 1.5 mL microcentrifuge tubes. To remove any residual liquid medium, pellets were re-suspended in phosphate-buffered saline (PBS) buffer and the cells collected by centrifugation at 12 300 x g for 15 minutes at room temperature. This procedure was repeated twice. Pellets were flash frozen by immersion into liquid nitrogen and stored at -80 °C until the proteomics analysis; frozen samples were transported on dry ice.

5.3.4 DPPH assay of extracts of isolated yeast cultures

Yeast cells were collected by centrifugation (5 minutes, 15 000 x g, room temperature) and re-suspended in 1 mL of cell lysis buffer composed of 50 mM tris–buffered saline (TBS) pH 7.6, mixed with 0.1 % (w/v) Triton X (Sigma Aldrich). The cells were disrupted with a sonicator probe (Model: VC250, Sonics & Materials Inc.), using a duty cycle of 40 % and an output of 3. Samples were kept on ice, and sonication was performed in 10 x 1-minute cycles, with 1 minute pause between the cycles. Cell lysis was confirmed by examination of the sample using a light microscope at magnification x400. The cell debris was removed by centrifugation (5 minutes, 15 000 x g, room temperature), and the supernatant was tested for free radical quenching antioxidant activity using the DPPH assay.

The antioxidant activity of extracts of UV-irradiated yeast samples and controls was measured using the colorimetric assay based on the neutralization of stable free radical 2,2-diphenyl-1-picrylhydrazyl (DPPH) (454). Briefly, 0.1 mL of each sample was mixed with 1.5 mL of 70 µM DPPH dissolved in methanol. The samples were shielded from light by aluminium foil and incubated for 30 minutes at room temperature, and the colour change from violet to yellow, when DPPH is reduced upon reaction with an antioxidant, was recorded at 515 nm using a UV/VIS spectrophotometer (model 7315, Jenway Ltd., Stone, Staffordshire, UK). A mixture of half-strength YPD medium (0.1 mL) and DPPH (1.5 mL) served as a control, and a mixture of methanol (0.1 mL) and DPPH (1.5 mL) served as the reaction blank. The percentage of DPPH radical scavenging activity was calculated as: $Scavenging\ activity\ (\%) = 100\ x\ (A_{blank} - A_{sample})/A_{blank}$. Experiments were performed in technical triplicates with three replicates, and scavenging activities were plotted as the mean of the 9 triplicate/replicate values against compound concentration.

5.3.5 Mass spectrometry analysis

The protein composition of the yeast samples was identified using Multidimensional Protein Identification Technology (MudPIT), with Tandem Mass Tags (TMT) used for relative quantification of labelled peptides, as follows:

Ten frozen yeast cell pellets (Sample 1 – 10, in duplicate) were processed for proteomics analysis. Cell pellets were homogenised in 100 µL of ice-cold lysis buffer using a micro-pestle. The lysis buffer consisted of 100 mM Tris(hydroxymethyl)aminomethane hydrochloride (pH 7.5), 300 mM NaCl, 2 % (v/v) Triton X-100, 1 % (w/v) sodium deoxycholate, 200 mM NaOH, 2 % (w/w) SDS, 2 % (v/v) β-mercaptoethanol, 2 x Complete Protease Inhibitor Cocktail (Roche) and 2 x Complete Phosphatase Inhibitor Cocktail (Roche). The cell pellets were homogenised for 30 seconds and placed on ice for 1 minute. The procedure was repeated four more times. Cell debris was pelleted by centrifugation (14 000 x g, 14 minutes, 4 °C) and the supernatants containing solubilised proteins were transferred to fresh 1.5 mL tubes and kept on ice.

For each yeast sample, the solubilised proteins were transferred into a fresh 2 mL microcentrifuge tube for protein precipitation. The solubilised proteins were combined with 800 µL of methanol, vortex agitated for 30 seconds and centrifuged at 14 000 x g for 1 minute at room temperature. Sample was mixed with 200 µL of chloroform, vortex agitated for 30 seconds, and centrifuged at 14 000 g for 1 minute at room temperature. 600 µL of deionised water was added to the mixture, and the mixture was vortex agitated for 30 seconds and

centrifuged at 14 000 x g for 10 minutes at room temperature. The upper aqueous phase of the sample was discarded, 800 µL of methanol were added to the remainder of the sample, and the precipitated proteins were collected by centrifugation at 14 000 x g for 5 minutes at room temperature. The supernatant was discarded and the pellet consisting of precipitated proteins was dried in Savant SpeedVac Concentrator (Thermo Fisher Scientific) for 1 minute, at room temperature. Precipitated proteins were solubilised in 150 µL of buffer composed of 100 mM Triethylammonium bicarbonate (TEAB) and 10 mM of Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (Sigma). Protein pellets were dissolved by 10 minutes of vortex agitation at room temperature. A 10 µL aliquot was taken for protein concentration determination by Bradford assay (Bio-Rad) using bovine serum albumin (Sigma) as a standard, and the remaining sample was incubated in the buffer (composed of 100 mM TEAB and 10 mM TCEP) for 1 hour at 55 °C to reduce protein disulphide bonds. For each yeast sample, a volume containing 30 µg of proteins (as calculated from Bradford assay) was transferred to a 1.5 mL tube and adjusted to 100 µL with buffer containing 100 mM TEAB and 10 mM TCEP (Sigma). Samples were alkylated by addition of 10 µL of buffer containing 100 mM TEAB, 10 mM TCEP and 198 mM iodoacetamide (Sigma), and incubated for 30 minutes at room temperature. The incubating samples were protected from light by aluminium foil. Proteins were digested by adding 10 µL of trypsin (Promega) solution (60 ng/µL) in 100 mM TEAB and incubating overnight at 37 °C, protected from light by aluminium foil.

Tandem Mass Tag (TMT) 10-plex reagents (Thermo Fisher Scientific TMT 10-plex kit, https://www.thermofisher.com/order/catalog/product/90110) were reconstituted according to the manufacturer's instructions, by adding 41 µL of acetonitrile (ACN) to 0.8 mg of each TMT label. The appropriate reconstituted label was added to the protein digests of yeast samples as follows:

Sample	**TMT Label Reagent**	**Sample**	**TMT Label Reagent**
1 (Control)	126	6 (2 hours of UV)	129N
2 (5 min of UV)	127N	7 (4 hours of UV)	129C
3 (10 min of UV)	127C	8 (8 hours of UV)	130N
4 (15 min of UV)	128N	9 (24 hours of UV)	130C
5 (1 hour of UV)	128C	10 (Control)	131

Samples were vortex agitated for 30 seconds and incubated at room temperature for one hour. The labelling reactions were quenched by the addition of 9 µL of 5 % (v/v) hydroxylamine

(Sigma) and incubated for 15 minutes. The 10 samples were combined in equal amounts of 10 µL. Excess labelling reagent was removed by solid phase extraction using a 1 mL Oasis HLB cartridge (Waters) as follows: The sample was prepared for purification in 4 % (v/v) ACN and 0.1 % (v/v) trifluoroacetic acid (TFA). A vacuum manifold was used to apply buffers and sample to the solid phase extraction cartridge at a rate 1 mL/minute. The solid phase extraction cartridge was conditioned with 2 x 1 mL of conditioning buffer consisting of 95 % (v/v) ACN and 0.1 % (v/v) TFA in deionised water. Conditioned cartridge was washed twice with 1 mL of wash buffer composed of 5 % (v/v) ACN and 0.1 % (v/v) TFA in deionised water. The sample was applied to the cartridge; the flow-through was collected and re-applied to the cartridge. Unbound contaminants were washed through with 5 x 1 mL of wash buffer, composed of 5 % (v/v) ACN and 0.1 % (v/v) TFA in deionised water, and the bound labelled peptides were eluted with 3 x 1 mL of elution buffer, composed of 85 % (v/v) ACN and 0.1 % (v/v) TFA in deionised water. The eluted labelled peptides were lyophilised in a SpeedVac (2 hours at room temperature).

The TMT labelled samples 1 - 10, containing lyophilised peptides labelled with TMT reagents, were reconstituted in 1.8 mL OFFGEL buffer consisting of 9.6 % (v/v) glycerol and 0.96 % (v/v) ampholytes in the form of IPG buffer pH 3-10 (GE Healthcare Life Sciences) in deionised water. The peptides of were solubilised using a sonic water bath for 10 seconds, followed by 30 minutes of vortex agitation at room temperature, and the insoluble material was removed by 15 minutes of centrifugation at 14 000 x g, at room temperature. The supernatant was applied to an isoelectric focusing (IEF) strip pH 3-10 (GE Healthcare Life Sciences) according to the manufacturer's instructions, and the labelled peptides were separated into 12 fractions and collected by an Agilent 3100 OFFGEL Fractionator. Isoelectric focusing was performed at 20 °C for a total of 20 kVh at a constant current of 50 µA. Once completed, the fractionation was held at 500 V until fraction collection. The fractions were collected in 1.5 mL tubes and acidified by the addition of TFA (final acid concentration of 0.1 % (v/v)). For each fraction, the salts, TFA and gel debris were removed by solid phase extraction using the procedure described above, with the exception that the fractionated peptides were eluted in 1 mL of buffer composed of 85 % (v/v) ACN and 0.1 % (v/v) formic acid in deionised water. Eluted peptides were lyophilised in a SpeedVac (2 hours at room temperature).

The peptides were solubilised in 50 mM ammonium bicarbonate for separation of the peptide mixture by liquid chromatography and analysis by tandem mass spectrometry. A portion of each fraction was analysed sequentially from fraction 1 (pH 3) to fraction 12 (pH 10). Chromatographic separations were performed using the Ultra-High Performance Liquid

Chromatography (UHPLC) system EASY-nLC II (Thermo Fisher Scientific). The peptides were separated using a reverse phase chromatography column 100 mm EASY-Column, with an internal diameter of 75 µm packed with a stationary phase of C18, 3 µm particles, and 120 Å porosity (Thermo Scientific). Peptides were eluted using a gradient of ACN (5 % to 40 % over 100 minutes, increased to 80 % over 10 minutes and held at 80 % for 5 minutes) and 0.1 % (v/v) formic acid. The flow rate of solvent was 300 nL / minute.

Mass spectra were acquired in the LTQ Orbitrap Velos Pro (Thermo Fisher Scientific) operated by Xcalibur™ software. The instrument was set to record the mass spectra ranging from 350 to 1800 m/z, at a resolution of 30,000. The 10 most intense precursor ions were subjected to sequencing by high-energy collision induced dissociation (CID) in the ion trap with a threshold of 5000 counts. The precursor ion selection isolation width was 2 units, and the normalised CID energy for precursor ion fragmentation was 35. Automatic gain control settings for FTMS survey scans were 105 counts and FT-MS/MS scans were 103 counts. Maximum acquisition time was 500 ms for survey scans and 250 ms for MS/MS scans. Charge-unassigned and single-charge state ions were excluded from the MS/MS analysis.

5.3.6 Data analysis: database searching

A database for spectra matching was constructed from the following sources:

- UniProt (www.uniprot.org) yeast protein sequences for Sporidiobolus salmonicolor and Rhodosporidium toruloides.
- Rhotordula minuta proteome (designation Rhomi1) acquired from The Fungal Genomics Resource (http://genome.jgi.doe.gov/programs/fungi/index.jsf).
- Sporobolomyces linderae CBS 7893 proteome (designation Spoli1) acquired from The Fungal Genomics Resource (http://genome.jgi.doe.gov/programs/fungi/index.jsf).
- Sporobolomyces roseus proteome (designation Sporo1) acquired from The Fungal Genomics Resource (http://genome.jgi.doe.gov/programs/fungi/index.jsf).
- Fungal basic leucine zipper sequences assembled using the BLASTp (483) search of the NCBI non-redundant (NR) database (340).
- Fungal proteins involved in biosynthesis of *Mycosporine-like amino acids*, assembled using the BLASTp (483) search of the NCBI non-redundant (NR) database (340).

Due to limitations of Mascot software, which does not support merging of databases that utilize different formats of protein sequence identifiers (such as UniProt and GenBank), each dataset

was processed independently and the results were merged after the database matching of tandem mass spectra.

Database matching of MS/MS spectra was performed using the Mascot software, version 2.2.03 (Matrix Science). Databases were installed in Mascot; Xcalibur raw files were processed into peak lists with Proteome Discoverer 1.4 (Thermo Fisher Scientific). The Proteome Discoverer Daemon was used to process the raw files with multidimensional protein identification technology (MudPIT) specifying up to 3 missed cleavages, a precursor ion mass tolerance of 20 ppm and a fragment ion tolerance of 0.8 Da. A variable/dynamic modification for oxidised methionine was set. Fixed/static modifications for carbamidomethylated cysteine and TMT-tagged lysine and N-termini were set. A target false discovery rate (FDR) for high confidence peptide hits was set to 0.01 (1 %) and a target FDR for medium confidence peptide hits was set to 0.05 (5 %). An independent search was conducted to assess labelling efficiency, by specifying all modifications as variable/dynamic. For all high and medium confidence peptides, 98 % were modified by TMT, 95 % of these were N-terminally labelled, and 96 % of lysines were modified by TMT labels.

5.3.7 Data analysis: quantification and result pre-processing

Detected proteins were grouped under the strict maximum parsimony principle. All detected peptides with a TMT modification of N-terminus were used to determine a normalisation factor for each label. All available reporter ion intensities were summed for each individual label, and the median of the summed intensities was determined for the ten labels. The normalisation factor for each label was obtained by expressing the median intensity over the sum of intensities, and applied to the raw reporter ion intensities for each respective label. Peptides for which all reported TMT reporter ions were detected and quantified were retained for further analysis. For each protein with multiple quantified peptide hits, the protein signal intensity was calculated as the mean value of TMT reporter ion intensities of peptides matched to the protein. The expression fold-changes of samples were calculated relative to Sample 1 (non-irradiated control).

All quantified proteins were annotated using InterProScan software (484) to predict the likely biological functions based on Gene Ontology (GO) terms (481) and Pfam profiles (485). Protein sequences were also annotated by BlastKOALA software (482) to assign KEGG pathways and KEGG modules (342). The results of computational annotation were manually curated by examining the primary literature and the information deposited in UniProt and NCBI protein databases. Proteins for which computational annotation was not successful, or resulted

in prediction of "predicted protein" or "unknown protein", were manually annotated by examining the results of BLASTp searches of UniProt and NCBI SwissProt databases, with E-value cut-off 0.001, and the function was assigned to the protein based on the related proteins if possible. Computational annotations and visualisation of results were conducted using *in-house* R scripts and python scripts.

5.3.7.1 Statistical analysis of annotated proteins with increased fold changes after UV-B exposure of yeast cells

Statistical analysis used the Fisher's exact test (486) to determine the GO terms over-represented (more frequent then would be expected based on the random distribution) in the dataset of proteins with significant (2-fold or greater) fold change in at least one UV-exposed sample. The test was conducted using the following procedure:

1) Quantified proteins were separated into *sensitive dataset (Ds)* and *control dataset (Dc)*. Proteins were included into *Ds* if protein fold changes in at least one UV-exposed yeast sample were 2.0 or higher. The remaining proteins were included into *Dc*.
2) For each GO term (*tested GO term*) from the list of all GO terms assigned to quantified proteins, the number of proteins was determined for:
   a) Proteins in *Ds* annotated with the *tested GO term* (*nsGO*)
   b) Proteins in *Dc* annotated with the *tested GO term* (*ncGO*)
   c) Proteins in *Ds* annotated with other GO terms (*nsO*)
   d) Proteins in *Dc* annotated with other GO terms (*ncO*)
3) One sided Fisher's exact test for over-representation of *tested GO term* in the *Ds* was performed on the following frequency distribution table:

	Tested GO	**Other GO**
Proteins in *Ds*	*nsGO*	*nsO*
Proteins in *Dc*	*ncGO*	*ncO*

The null hypothesis (H_0) of the test was that the *tested GO term* is equally frequent amongst the proteins in dataset *Dc* and amongst the proteins in dataset *Ds*, while the alternative hypothesis (H_A) was that the tested GO is over-represented amongst the proteins from *Ds*. The *tested GO term* was considered significantly over-represented for Fisher's exact test p-value <= 0.05. To assess the robustness of this approach, equivalent Fisher's exact tests were also performed where proteins were included into *Ds* dataset if the protein fold-changes were 1.5 or higher, and if the fold-changes were 4.0 or higher. In the results presented, the GO terms were considered

significantly over-represented if p-values of at least 2 out of these 3 tests were below the 0.05 threshold, or if p-value was below 0.05 in the fold-change >= 4.0 test.

Statistical analysis of over-representation of KEGG Pathways and KEGG Modules was performed using the methodology for statistical analysis of GO terms (described previously), with the exception that the GO terms were replaced with the KEGG Pathway terms or with the KEGG Modules during the step 2) of the statistical analysis.

5.3.7.2 Statistical analysis of annotated proteins with fold change reduction after UV-B exposure of yeast cells

The statistical analysis was conducted to determine the GO terms, KEGG pathways and KEGG modules significantly over-represented in the dataset of proteins showing a fold change reduction in UV-B exposed yeast cultures. This analysis was conducted using the methodology described in 5.3.7.1, with the exception that the proteins were included into the *sensitive dataset (Ds)* if the proteins showed the fold change reduction of 2.0 or higher in at least one UV-exposed yeast culture.

5.4 Results

5.4.1 Isolation of UV-tolerant yeast samples

Twelve yeasts isolated from the leaves of various Brazilian plants (designated Y1 to Y12), five yeast samples previously isolated by Casteliani et al. (LEV-2, LEV-9, LEV-12, LEV-13 and LEV-16) (479), and the baker yeast *Saccharomyces cerevisiae* were tested for viability after the exposure to UV-B irradiation. The baker yeast was used as a control based on the study by Pulschen et al. (2015), who used it as UV-sensitive yeast control to evaluate the viability of UV-resistant yeasts from high-altitude extreme environments (478).

The survival rates of yeasts Y1 – Y12 exposed to UV-B irradiation are listed in Table 5.1. The UV-tolerance of these yeasts was no higher than the baker yeast *S. cerevisiae* control, and these isolates were not considered for further study.

Table 5.1 Survival rates of UV-B irradiated yeast isolates Y1 – Y12

Table lists the survival rates of yeast isolates Y1 to Y12, and the baker yeast *S. cerevisiae* (By) exposed to long-term UV-B radiation. The survival rates were determined as ratios of colony forming units (CFUs) of UV-irradiated yeast samples to CFUs of non-irradiated controls. Colonies were counted after 48 hours of yeast growth. The numbers listed are the mean values of three technical triplicates; errors are expressed as one standard deviation of the mean.

Yeast sample	**Survival rate after the exposure to:**		
	1 hour of UV-B	**8 hours of UV-B**	**24 hours of UV-B**
Y1	40 % ± 4 %	16 % ± 5 %	0 %
Y2	25 % ± 6 %	10 % ± 1 %	2 ± 1 %
Y3	17 % ± 2 %	0 %	0 %
Y4	37 % ± 3 %	16 % ± 8 %	0 %
Y5	28 % ± 3 %	13 % ± 4 %	0 %
Y6	29 % ± 8 %	15 % ± 6 %	0 %
Y7	22 % ± 2 %	1 % ± 1 %	0 %
Y8	10 % ± 3 %	8 % ± 4 %	0 %
Y9	17 % ± 6 %	0 %	0 %
Y10	24 % ± 3 %	2 % ± 2 %	0 %
Y11	35 % ± 3 %	10 % ± 3 %	0 %
Y12	22 % ± 11 %	0 %	0 %
By	27 % ± 4 %	13 % ± 5 %	0 %

The survival curves of yeast samples LEV-2, LEV-9, LEV-12, LEV-13 and LEV-16 are presented in Figure 5.1. UV-B exposure for 24 hours was lethal to the baker yeast *S. cerevisiae*, while the yeasts LEV-12 and LEV-16 showed survival rates below 10 %. The isolates LEV-2, LEV-9 and LEV-13 showed survival rates above 60 %. The isolate LEV-2 had the highest rate of survival amongst the tested yeasts with a survival rate of ~80 % after 8 hours of UV-B exposure and ~75 % after 24 hours of UV-B exposure.

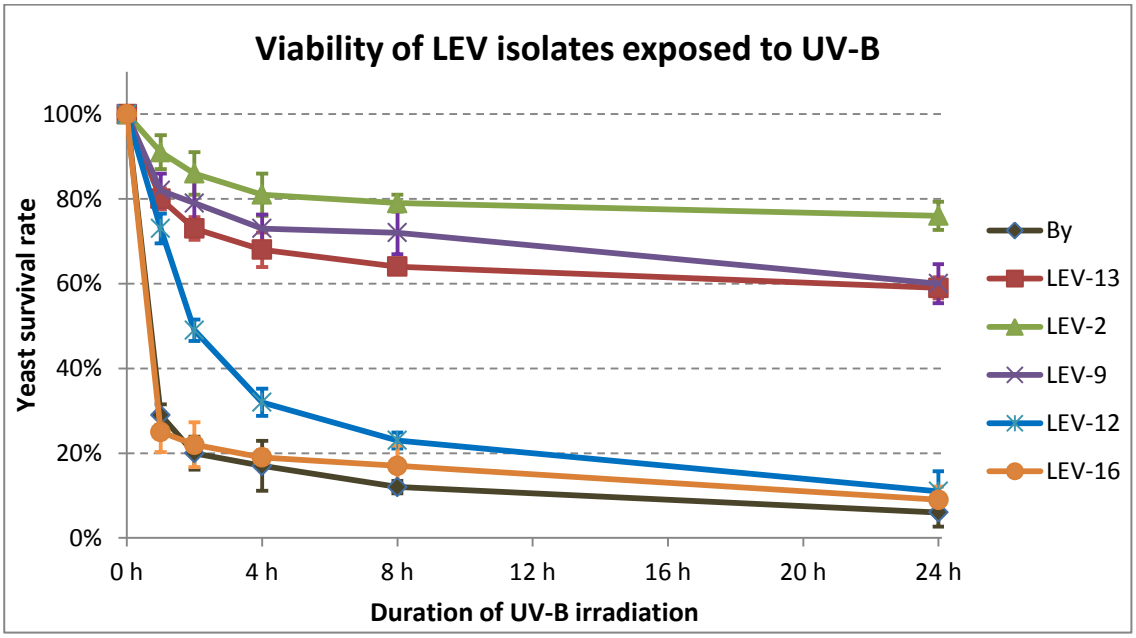


Figure 5.1 Survival rate of LEV yeast isolates upon exposure to UV-B

The figure displays the survival rates of UV-B exposed yeast isolates (LEV-2, LEV-9, LEV-12, LEV-13 and LEV-16) and baker yeast *S. cerevisiae* (By), colour coded as indicated on the chart. The survival rates were determined as ratios of CFUs of UV-irradiated yeast samples to CFUs of non-irradiated controls. Colonies were counted after 48 hours of yeast growth. Points represent mean values of three technical replicates, and error bars indicate one standard deviation of the mean.

The yeasts LEV-2, LEV-13 and LEV-9, which showed the high rate of survival after 24 hour exposure to UV-B, were next evaluated for antioxidant activity. The extracts of these yeasts, prepared by cell lysis and removal of insoluble material, were tested using the colorimetric assay based on the quenching of 2,2-diphenyl-1-picrylhydrazyl (DPPH), stable free radical (Figure 5.2). Extracts of all three tested yeasts showed an increase in DPPH quenching activity after 24 hours of UV-B irradiation. The increase in antioxidant activity was moderate (~25 % increase) for samples LEV-9 and LEV-13, while the extract of yeast LEV-2 showed ~75 % increase in antioxidant activity after 24 hours of UV-B irradiation. The yeast LEV-2 also showed ~50 % reduction in DPPH quenching for samples exposed to 1 hour and 2 hours of UV-B.

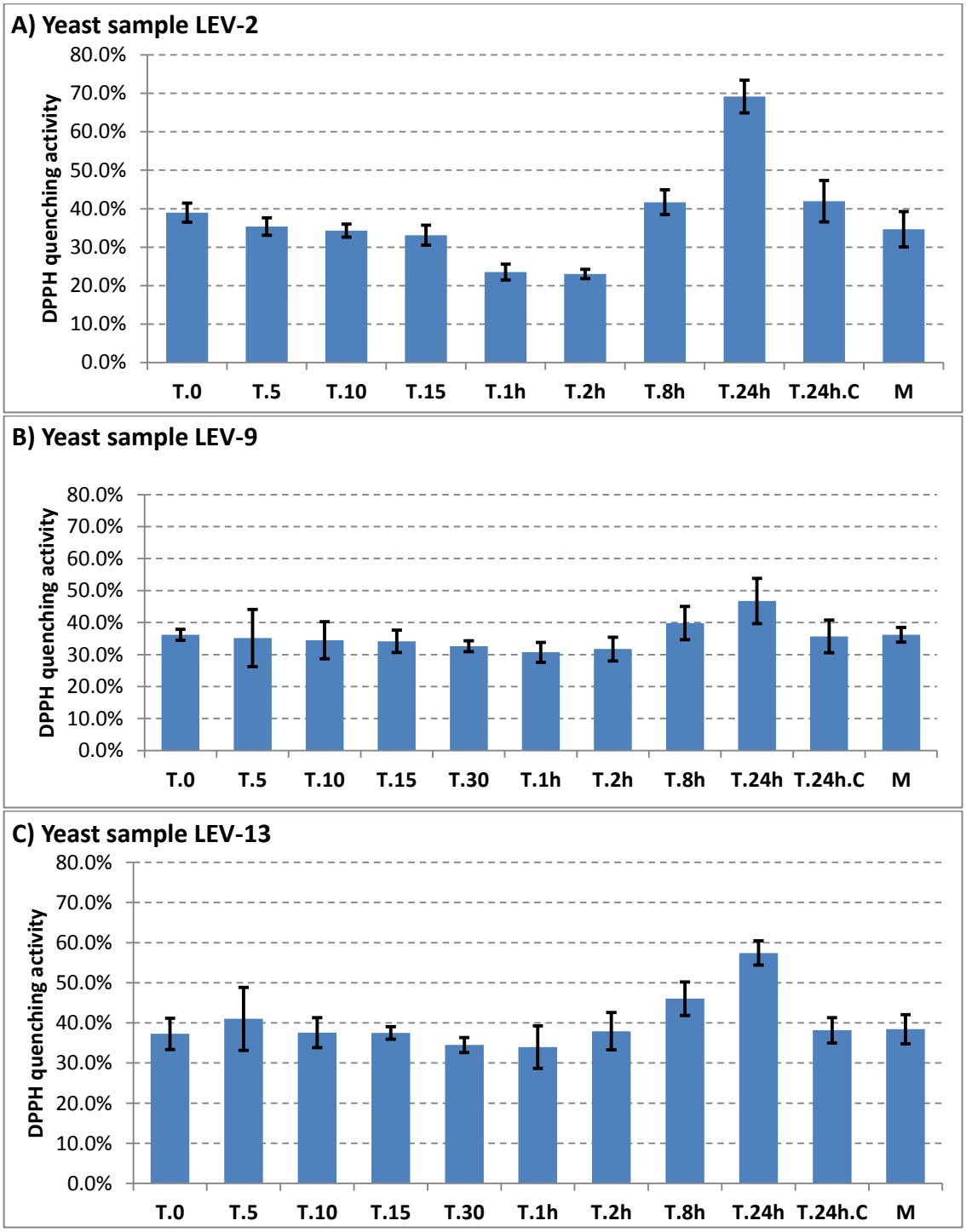


Figure 5.2 DPPH quenching activity of LEV yeast isolates exposed to UV-B

The figure displays the DPPH free radical quenching activity of extracts of yeast isolates LEV-2 (A), LEV-9 (B) and LEV-13 (C). Yeast cultures were irradiated for time periods ranging from 5 minutes (sample T.5) to 24 hours (sample T.24), using dual Philips Ultraviolet-B TL 20W/12RS lamps, and yeast extracts were obtained by cell lysis and removal of insoluble material. The controls were half-strength YPD medium (sample M), non-irradiated yeast cultures (sample T.0), and non-irradiated yeast cultures grown for 24 hours (sample T.24.C). Error bars indicate 1 standard deviation of the mean, calculated from the three experiments; DPPH quenching values of three technical triplicates of each experiment were averaged.

The extract of yeast LEV-2, previously identified as *Sporobolomyces sp.* (479), was chosen as the model organism for further study because of its high tolerance to UV-B (Figure 5.1) and the observed induction of DPPH free radical quenching activity (Figure 5.2). To assess if this yeast adapts to UV-induced stress during the UV-B exposure, the rate of yeast cell death was calculated for different time periods of UV-B exposure. The results presented in Figure 5.3 indicate that the initial two hours of UV-B exposure cause moderate loss of viability in yeast LEV-2 (~9% during first hour, followed by ~6% reduction in viability in remaining yeasts in second hour), but the rate of cell death diminishes over the following time periods, and the cell death rate of LEV-2 yeast culture is ~0.6% of viable colony forming units per hour after 8 hours of UV-B exposure.

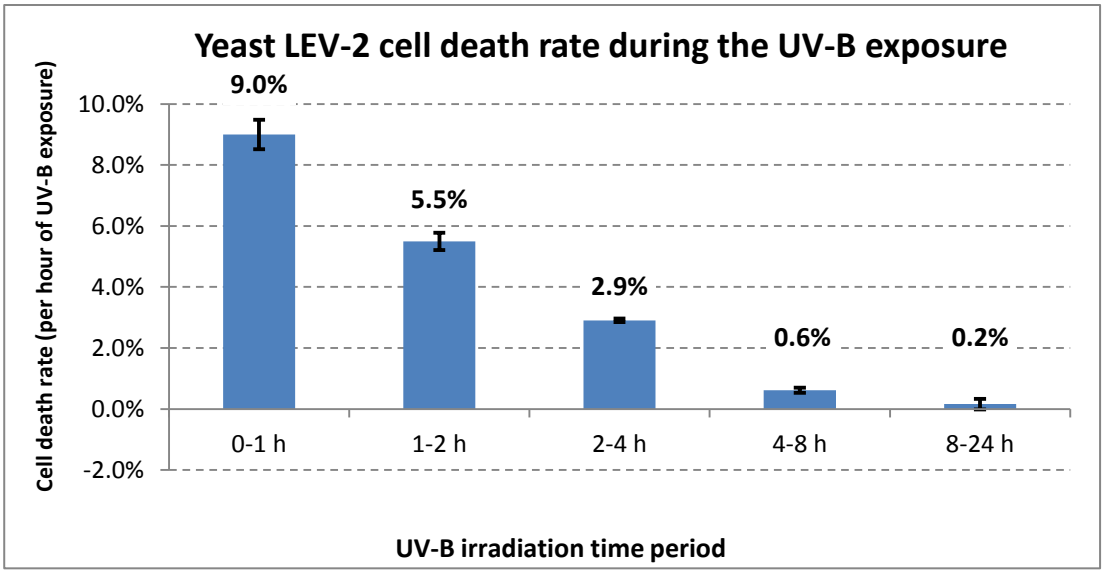


Figure 5.3 Death rate of yeast LEV-2 exposed to UV-B radiation

Figure displays the rate of cell death of LEV-2 yeast irradiated using dual Philips Ultraviolet-B TL 20W/12RS lamps for each of listed time periods. The cell death rate values are denoted above bars, and are expressed as percentage reduction in number of viable colony forming units per hour of UV-B irradiation. The error bars represent one standard deviation of the mean of three experimental replicates.

5.4.2 MudPIT analysis of UV-tolerant yeast LEV-2

The *Sporobolomyces* yeast LEV-2 was grown in liquid medium and exposed to increasing duration of UV radiation. The irradiation was performed for time intervals ranging from 5 minutes to 24 hours, with UV output of 4 J/m^2/s UV-B and 1.75 J/m^2/s UV-A. The solubilised proteins were extracted from UV-irradiated samples and non-irradiated controls, labelled by TMT chemical tags, and quantified by multidimensional protein identification technology (MudPIT). The analysis identified 751 proteins for which fold changes could be determined (Appendix C-2). Based on the previously published proteomics studies (487–489), protein expression fold changes of 2 or higher were considered significant.

227 proteins (~30% of quantified proteins) showed a significant fold change increase (2 or higher) in irradiated yeast LEV-2 cultures (Figure 5.4/A). The median value of fold changes of these 227 proteins was ~1 for LEV-2 controls, and for samples exposed to UV-B for up to 4 hours; ~1.5 for LEV-2 exposed to 8 hours of UV-B; and ~2.5 for LEV-2 exposed to 24 hours of UV-B. 279 proteins (~37% of quantified proteins) showed significant fold change decrease in UV-irradiated yeast cultures (Figure 5.4/B). For these 279 proteins, the median fold change value was ~1.0 for non-irradiated control and for LEV-2 exposed to 5, 10 and 15 minutes of UV, ~0.6 for LEV-2 culture irradiated for 1 hour or 8 hours, ~0.4 for LEV-2 exposed to 2 hours or 4 hours of UV, and ~0.8 for yeast culture irradiated for 24 hours and for non-irradiated yeast culture grown for 24 hours.

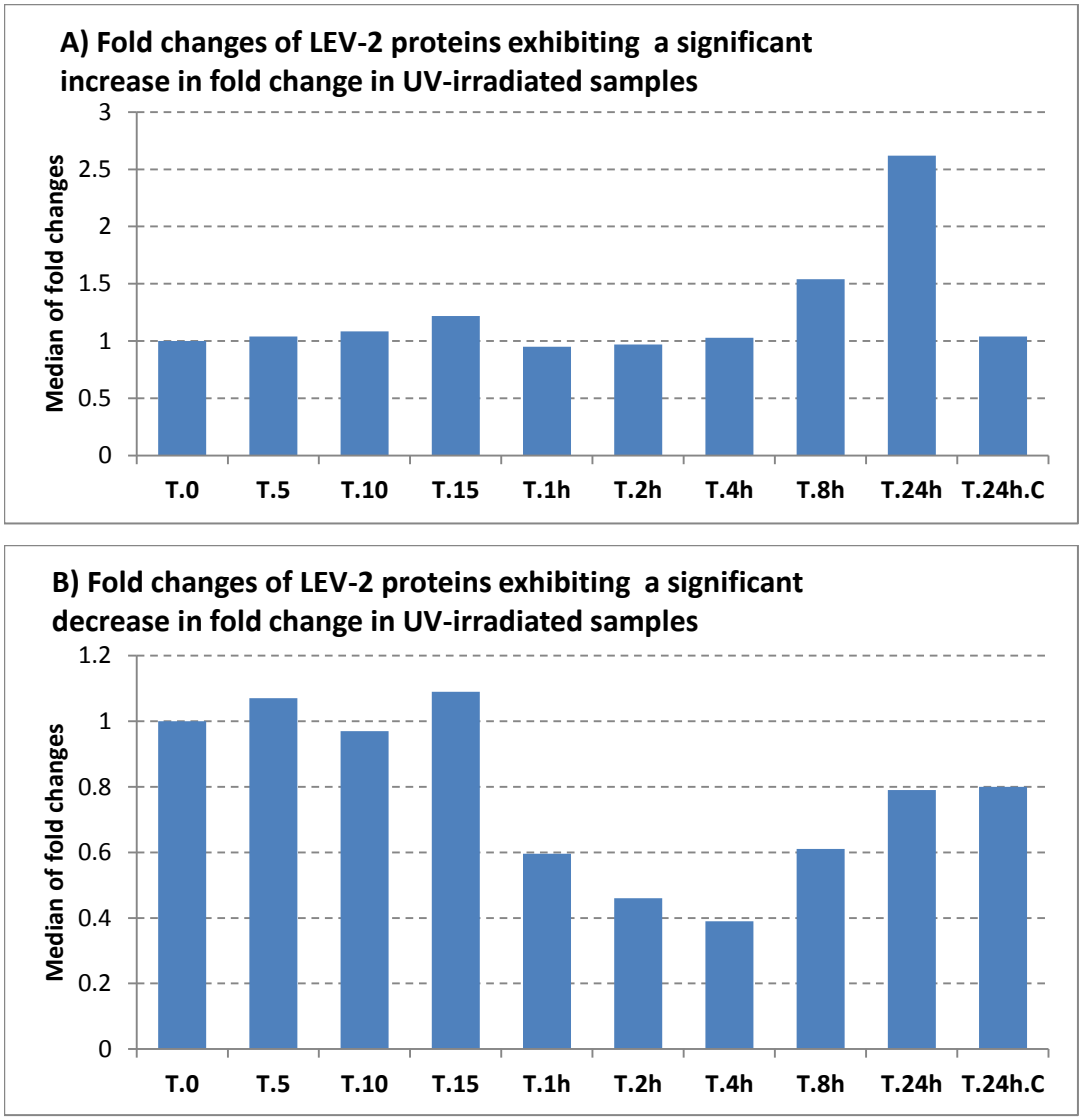


Figure 5.4 Expression profiles of proteins exhibiting a significant fold change

The figure (A) displays the median values of fold changes for 227 yeast LEV-2 proteins that exhibited 2-fold or higher increase in expression in at least one UV-irradiated sample, relative to the non-irradiated control. The figure (B) shows the median values of expression fold changes for 279 yeast LEV-2 proteins that exhibited significant (2-fold) or higher decrease in fold change in at least one UV-irradiated sample.

5.4.3 Functional annotation of yeast LEV-2 proteome

The 752 yeast LEV-2 proteins, quantified by MudPIT analysis, were annotated for predicted biological function using the InterProScan (484) tool to assign Gene Ontology (GO) terms (481) to the proteins. In addition, KEGG tools (482) were used to assign the predictions of biochemical pathways (KEGG pathways) and biological functions (KEGG Modules) to the quantified proteins. Statistical analysis based on Fisher's exact test (486), described in section 5.3.7.1, was performed to identify GO terms, KEGG modules and KEGG pathways over-represented amongst the proteins showing fold change increases and fold change reductions in LEV-2 cultures exposed to UV-B radiation.

For the dataset of 227 proteins exhibiting a fold change increase in LEV-2 isolate exposed to UV-B, the over-represented GO terms (Table 5.2/A) included GOs related to cellular transport systems, ribosome biogenesis, stress response, cellular signalling and cellular respiration. The annotations based on the KEGG pathways (Table 5.3/A) and KEGG modules (Table 5.4/A) also indicated that the functions and pathways related to stress response, cellular signalling and cellular respiration are over-represented in this dataset. In addition, the over-represented KEGG pathways also included pathways involved in metabolism of arginine, histidine and mannose.

279 LEV-2 proteins exhibited a fold change reduction in UV-exposed yeast cultures. The GO terms over-represented amongst these proteins (Table 5.2/B) included GO terms related to protein biosynthesis, protein folding and degradation, ATP binding and synthesis, pentose-phosphate pathway, biosynthesis of nucleotides and metabolism of certain amino acids such as glycine and serine. The annotation by KEGG pathways (Table 5.3/B) and KEGG modules (Table 5.4/B) identified that the pathways involved in carbohydrate metabolism and ribonucleotide biosynthesis are over-represented in this dataset.

Table 5.2 GO terms over-represented in datasets of LEV-2 proteins showing a significant fold in UV-B exposed yeast cultures

The column A lists GO terms over-represented in a dataset of proteins exhibiting a significant fold change increase in UV-B exposed yeast LEV-2 samples, while the column B lists GO terms over-represented amongst proteins showing a significant fold change reduction. Terms marked by a star (*) were over-represented in a dataset of proteins with expression fold change of 4.0. The GO terms were considered over-represented if Fisher's exact test resulted in p-value below 0.05, when frequencies of terms were compared between all proteins and proteins with fold-changes 2.0 or higher.

A) GO terms over-represented amongst proteins with fold change increase in UV-B exposed samples	**B) GO terms over-represented amongst proteins with fold change decrease in UV-B exposed samples**
ATP phosphoribosyltransferase activity (GO:0003879)	3-deoxy-7-phosphoheptulonate synthase activity (GO:0003849)*
Cytochrome-c oxidase activity (GO:0004129)	ATP binding (GO:0005524)
Histidine biosynthetic process (GO:0000105)*	ATP-dependent peptidase activity (GO:0004176)
Intracellular (GO:0005622)	Cytidylate kinase activity (GO:0004127)
Mitochondrial inner membrane (GO:0005743)	Cytoplasm (GO:0005737)
Polyamine biosynthetic process (GO:0006596)	*De novo* pyrimidine nucleobase biosynthetic process (GO:0006207)*
Response to stress (GO:0006950)	FK506 binding (GO:0005528)*
Ribosome (GO:0005840)	Glycine hydroxymethyltransferase activity (GO:0004372)*
Small GTPase mediated signal transduction (GO:0007264)	Glycine metabolic process (GO:0006544)*
Structural constituent of ribosome (GO:0003735)	Heme binding (GO:0020037)
Translation (GO:0006412)	Histone peptidyl-prolyl isomerization (GO:0000412)*
Translational elongation (GO:0006414)	L-serine metabolic process (GO:0006563)*
Transmembrane transport (GO:0055085)	Misfolded or incompletely synthesized protein catabolic process (GO:0006515)
Transport (GO:0006810)*	Nucleobase-containing compound kinase activity (GO:0019205)
Transporter activity (GO:0005215)	Nucleobase-containing compound metabolic process (GO:0006139)
Unfolded protein binding (GO:0051082)*	Pentose-phosphate shunt (GO:0006098)
	Peptidyl-proline modification (GO:0018208)*
	Protein binding (GO:0005515)*
	Proteolysis (GO:0006508)*
	Protein folding (GO:0006457)
	Protein refolding (GO:0042026)
	Pyrimidine nucleotide biosynthetic process (GO:0006221)
	Serine-type endopeptidase activity (GO:0004252)
	Transferase activity (GO:0016740)*
	Translation initiation factor activity (GO:0003743)
	Translational initiation (GO:0006413)*
	Uridylate kinase activity (GO:0009041)

Table 5.3 KEGG pathways over-represented amongst the LEV-2 proteins showing a significant fold change in yeast LEV-2 exposed to UV-B

The column A lists KEGG pathways over-represented in a dataset of proteins exhibiting a significant fold change increase in UV-B exposed yeast LEV-2 cultures, while the column B lists pathways over-represented amongst proteins showing a significant fold change decrease. The pathways were assigned by BlastKOALA search followed by KEGG pathway analysis. KEGG pathways marked by a star (*) were over-represented in a dataset of proteins with expression fold change of 4.0 or higher. The KEGG pathways were considered over-represented if Fisher's exact test resulted in p-value below 0.05, when frequencies of KEGG pathway terms were compared between all proteins and proteins with fold-changes 2.0 or higher.

A) KEGG pathways over-represented amongst the proteins with a fold change increase in UV-B exposed LEV-2	**B) KEGG pathways over-represented amongst the proteins with a fold change decrease in UV-B exposed LEV-2**
Arginine and proline metabolism*	Cyanoamino acid metabolism*
Calcium signaling pathway*	HIF-1 signaling pathway
cAMP signaling pathway	One carbon pool by folate*
cGMP-PKG signaling pathway	Pyrimidine metabolism*
Fructose and mannose metabolism	
Glutathione metabolism*	
Histidine metabolism	
Pentose and glucuronate interconversions	
PI3K-Akt signaling pathway*	
Ras signaling pathway	
Ribosome	

Table 5.4 KEGG modules over-represented amongst the LEV-2 proteins showing a significant fold change in yeast LEV-2 exposed to UV-B

The column A lists KEGG modules over-represented in a dataset of proteins exhibiting a significant fold change increase in UV-B exposed yeast LEV-2 samples, and the column B shows KEGG modules over-represented in a dataset of proteins showing a significant fold change decrease. The KEGG modules were assigned by BlastKOALA search followed by KEGG module analysis. KEGG modules marked by a star (*) were over-represented in a dataset of proteins with expression fold change of 4.0 or higher. The KEGG modules were considered over-represented if Fisher's exact test resulted in p-value below 0.05, when frequencies of KEGG module terms were compared between all proteins and proteins with fold-changes 2.0 or higher.

A) KEGG modules over-represented amongst the proteins with a fold change increase in UV-B exposed LEV-2	**B) KEGG modules over-represented amongst the proteins with a fold change decrease in UV-B exposed LEV-2**
Cytochrome c oxidase	C1-unit interconversion
	Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate
Polyamine biosynthesis, arginine => ornithine => putrescine*	
Ribosome, eukaryotes	Pentose phosphate pathway (Pentose phosphate cycle)
	Pyrimidine ribonucleotide biosynthesis, UMP => UDP/UTP,CDP/CTP*

5.4.4 Yeast LEV-2 proteins involved in response to UV-B induced stress

The functional annotation of 751 quantified yeast LEV-2 proteins identified 105 proteins involved in cellular stress responses. Of these, four proteins were annotated as basic leucine zipper proteins, 22 as cellular signalling proteins, 17 as enzymes involved in biosynthesis of small molecule antioxidants, 22 as enzymatic antioxidants, six as enzymes involved in DNA repair, 24 as heat shock proteins, and 10 proteins were annotated as enzymes involved in the biosynthesis of mycosporine-like amino acids (MAAs).

5.4.4.1 Basic-leucine zipper proteins

Four basic leucine zipper (bZip) containing proteins, similar to human Nrf2 and AP-1 proteins, were identified and quantified by MudPIT analysis using the database of fungal bZip proteins (Figure 5.5). The protein similar to XP_007274754.1 (bZip transcription factor of fungi *C. gloeosporioides*) exhibited a significant fold change increase in LEV-2 yeast cultures exposed to 5 minutes, 1 hour, 2 hour and 4 hour of UV-B. The proteins similar to GAP83664.1 (bZip protein from *R. necatrix*) and XP_748177.1 (bZip protein from *A. fumigatus*) showed a significant fold change decrease in the yeast cultures exposed to 4 hours and 8 hours of UV-B, and the bZip protein KFA50940.1 showed no significant fold changes in UV-B exposed cultures.

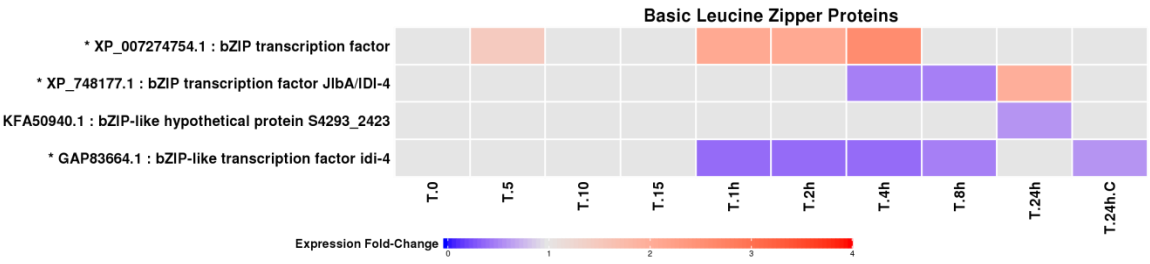


Figure 5.5 Fold change profiles of yeast LEV-2 bZip proteins

Figure displays a heat-map of yeast isolate LEV-2 bZip protein fold changes in cultures exposed to UV-B for 5 minutes (T.5) to 24 hours (T.24h) and non-irradiated controls (T.0 and T.24h.C). The proteins with fold change reduction in the UV-exposed samples are coloured blue, while proteins with fold change increases are coloured red. Proteins showing a fold change lower than 1.5 are coloured light-grey. The protein identifiers of the proteins showing 2-fold or higher fold change in at least one UV-exposed yeast culture are also marked with a star (*).

5.4.4.2 Signalling proteins

The quantitative MudPIT analysis identified 22 signalling proteins (Figure 5.6). These were annotated as proteins involved in various signalling pathways including FoxO signalling (490), MAPK signalling (491) and RAS signalling (471). Three 14-3-3 proteins were quantified, and all three exhibited fold change reduction in LEV-2 cultures irradiated for 1 hour to 4 hours. Four MAPK-signalling kinases were quantified, and all four exhibited fold-change increase in yeasts irradiated for 8 hours and 24 hours, while two MAPK-signalling kinases (similar to M6XZ23 and Spoli1_184897 proteins) also showed moderate (~1.5-fold or lower) reduction in expression in LEV-2 cultures irradiated for 1 hour to 4 hours. Annotation identified two proteins involved in FoxO signalling, both of which exhibited moderate fold change increase in LEV-2 exposed to 8 hours and 24 hours of UV-B. Two Ras-related proteins were quantified; both of these proteins showed fold change increase in LEV-2 exposed to 24 hours of UV-B, while the Ras-related protein M7WXY7 also showed moderate, ~1.5, fold change increase in yeast cultures exposed to 15 minutes of UV-B. Four cell division control (Cdc) proteins were quantified; of these, two proteins belonging to Cdc42 family showed fold change increase in yeasts exposed to 8 hours and 48 hours of UV-B, while two Cdc48 proteins exhibited fold change reduction in LEV-2 irradiated for 2 hours and 4 hours. A single Hippo-signalling protein was identified in this study; this protein did not show significant fold changes in UV-B irradiated LEV-2 cultures or in the controls. One calcium signalling protein calmodulin was identified and quantified, and showed a fold change increase in samples exposed to 1 hour to 4 hours of UV-B. The MudPIT analysis also identified three adenylate kinases and two other kinases for which detailed annotation could not be determined. The adenylate kinases showed fold change decrease in LEV-2 exposed to 1 hour to 24 hours of UV-B, the kinase Rhomi1_185026 exhibited fold change increase in yeast irradiated for 24 hours, and the kinase Rhomi1_141225 showed moderate fold change increase in LEV-2 irradiated for 5 to 15 minutes, 8 hours and 24 hours.

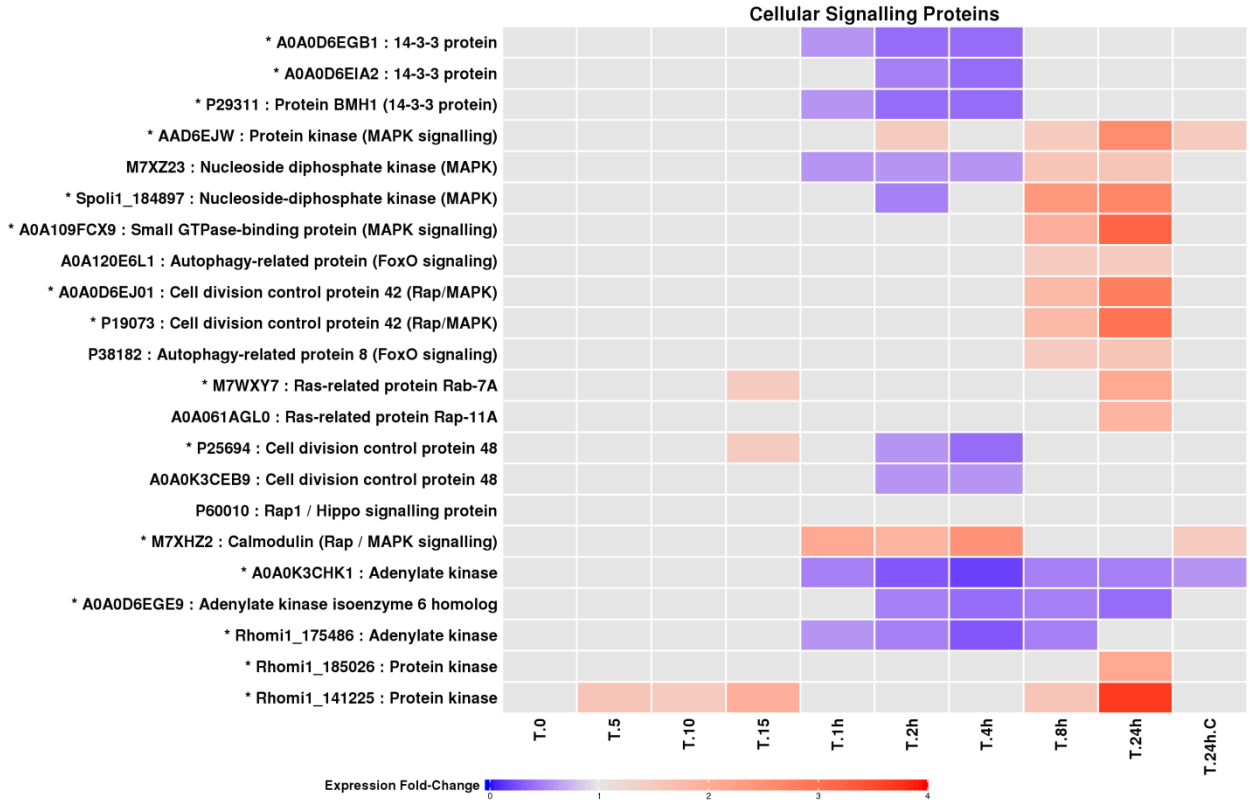


Figure 5.6 Fold change profiles of LEV-2 proteins involved in cellular signalling

Figure displays a heat-map of fold changes of LEV-2 proteins involved in cellular signalling, cell cycle control and apoptosis. The fold changes are shown for LEV-2 cultures exposed to UV for 5 minutes (T.5) to 24 hours (T.24h) and in the cultures of non-irradiated controls (T.0 and T.24h.C). The proteins showing a fold change reduction in the UV-exposed samples are coloured blue, while proteins with fold change increase are coloured red. Proteins showing a fold change lower than 1.5 are coloured light-grey, and identifiers of proteins showing 2-fold or higher fold change in at least one UV-exposed yeast culture are also marked with a star (*).

5.4.4.3 Proteins involved in biosynthesis of antioxidants

MudPIT analysis of yeast isolate LEV-2 exposed to UV-B irradiation identified and quantified 17 proteins involved in biosynthesis of small-molecule antioxidants such as glutathione and ubiquinol (Figure 5.7). The hydroxyacylglutathione hydrolase (Rhomi1_37826), protein involved in biosynthesis of glutathione (492), showed fold change reduction in LEV-2 cultures exposed to 4 hours to 24 hours of UV-B. Four glutathione transferase (GST) enzymes, which mediate the phase II detoxification by catalysing the conjugation of xenobiotic functional group with glutathione (492), were identified. The GST protein M7WWF5 exhibited fold change reduction in samples irradiated for 1 hour to 24 hours, while GSTs Rhomi1_153124 and Rhomi1_19319 showed moderate fold change decrease in yeasts irradiated for 2 hours and 4 hours. The GST protein Rhomi_153124 also showed a fold change increase in LEV-2 exposed to 15 minutes and 24 hours of UV-B. The PdxS/SNZ family lyases and pyridoxine 4-dehydrogenases are involved in yeast biosynthesis of vitamin B6, and are associated with resistance to oxidative stress (493). Four lyases of PdxS/SNZ protein family were quantified by MudPIT analysis; these proteins showed fold change decrease in UV-irradiated samples (1 hour to 8 hours of UV-B). Single pyridoxine 4-dehydrogenase (Rhomi1_173574) was identified and quantified, and this protein showed no change in expression levels in UV-irradiated yeast cultures. Eight succinate dehydrogenase enzymes, involved in reduction of oxidised coenzyme Q10 to its reduced form (ubiquinol) (494), were identified and quantified in this study. Of these, protein P47052 showed no significant fold changes in any of the studied samples; A0A109FAY5 and Sporo_19407 exhibited moderate, ~1.5, fold change decrease in yeast LEV-2 exposed to 4 hours of UV-B; Spoli1_173565 exhibited moderate fold change decrease in samples exposed for 1 hour, 4 hours and 8 hours; Rhomi1_182330 showed moderate fold change decrease in yeast irradiated for 2 hours and moderate fold change increase in sample exposed to 24 hours; and Sporo1_9115 and Rhomi1_167063 exhibited a fold change increase in yeast culture irradiated for 24 hours.

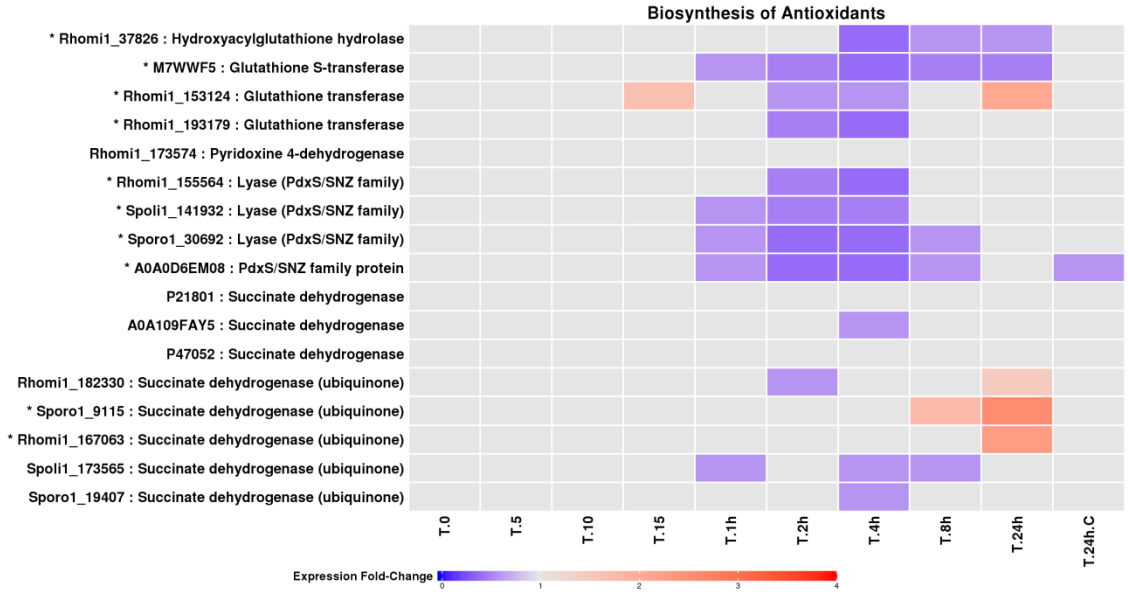


Figure 5.7 Expression profiles of LEV-2 enzymes involved in biosynthesis of antioxidants

Figure displays the heat-map of the fold changes of yeast LEV-2 enzymes involved in the metabolism of small-molecule antioxidants. The LEV-2 cultures were exposed to UV for 5 minutes (T.5) to 24 hours (T.24h), and compared to non-irradiated controls (T.0 and T.24h.C). The proteins showing a fold change reduction in the UV-exposed samples are coloured blue, while proteins showing a fold change increase are coloured red. The proteins exhibiting fold-changes lower than 1.5 are coloured light-grey, and the proteins showing 2-fold or higher fold change in at least one UV-exposed sample are marked with a star (*).

5.4.4.4 Enzymatic antioxidants

The MudPIT study of UV-exposed yeast LEV-2 identified 22 enzymatic antioxidants, of which 6 showed a significant fold change increase and 9 exhibited a significant fold change decrease in yeast LEV-2 cultures exposed to UV-B (Figure 5.8). The identified antioxidants included superoxide dismutase (SOD), catalase, aldehyde dehydrogenase (ALDH), glutathione peroxidase (GTRx), isocitrate dehydrogenase (IDH) and cytochrome-c peroxidase (CCP) proteins, all of which are major enzymatic antioxidants in yeasts as well as animals (495–500).

One ADH and one benzaldehyde dehydrogenase enzyme were identified and quantified in this study, and both enzymes showed significant fold change increase in LEV-2 isolates exposed to 24 hours of UV-B; the ADH also exhibited moderate, ~1.5, fold change increase in yeast exposed to 8 hours of UV-B. The single catalase was quantified; this enzyme showed moderate reduction in expression levels (~1.5 fold change decrease) in LEV-2 cultures exposed to 2 hours to 8 hours of UV-B. Five CCP enzymes were quantified in yeast samples; two of these (Sporo1_16456 and A0A0D6ERS5) showed the fold change reduction in LEV-2 cultures

exposed to 1 hour to 24 hours of UV-B; CCP Sporo_190216 exhibited a fold change increase in samples exposed to 15 minutes, 8 hours and 24 hours of UV-B and fold change decrease in samples exposed to 1 hour and 2 hours of UV; CCP Rhomi1_42978 showed a fold change decrease in LEV-2 cultures exposed to 1 hour to 4 hours of UV and moderate fold change increase in culture irradiated for 24 hours; and CCP Rhomi1_149378 showed a fold change decrease in samples irradiated for 2 hours to 8 hours, and an increase in expression in LEV-2 culture exposed to 15 minutes of UV-B. Two GTRx enzymes were identified by MudPIT proteomics analysis; the GTRx enzyme P38143 exhibited fold change reduction in yeasts exposed to 1 hour and 2 hours of UV, while the other GTRx (Rhomi1_146413) showed no changes in expression levels in UV-exposed yeast samples. The proteomics analysis identified two NAD+ dependant IDHs, both of which exhibited increase in fold changes in samples subjected to high dose of UV (8 to 24 hours of UV), and four NADP+ associated IDHs which displayed various degrees of fold change reduction in LEV-2 isolates exposed to 1 hour or longer UV-irradiation. Four SOD enzymes were identified and quantified, and showed significant fold change increase in LEV-2 samples irradiated for 24 hours. The SODs Rhomi1_168331 and Spoli1_20960 also exhibited fold-change increase in yeast exposed to 8 hours, and the SOD Rhomi1_86056 showed increase in fold changes in LEV-2 irradiated for 5 to 15 minutes and for 1 hour.

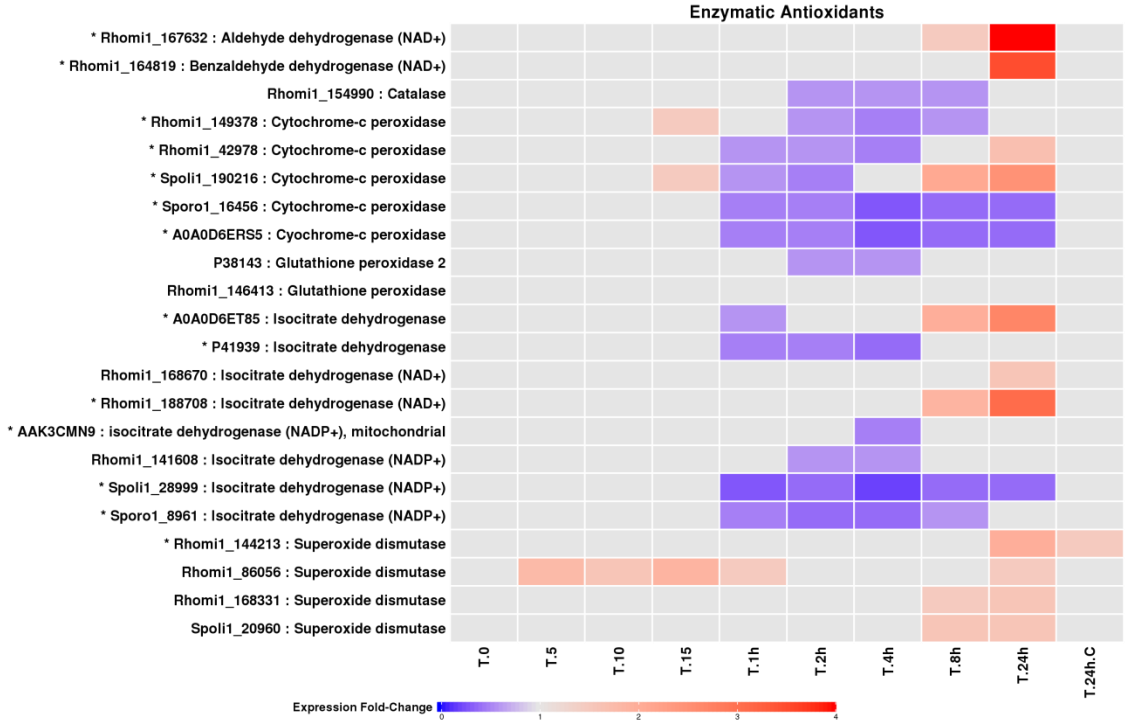


Figure 5.8 Fold change profiles of LEV-2 enzymatic antioxidants

The heat-map shows fold changes of enzymatic antioxidants of yeast LEV-2. The yeast cultures were exposed to UV for 5 minutes (T.5) to 24 hours (T.24h), and non-irradiated cultures were used as controls (T.0 and T.24h.C). The proteins showing fold change reduction in the UV-exposed samples are coloured blue, while proteins with fold change increase are coloured red. The proteins exhibiting fold change lower than 1.5 are coloured light-grey, and identifiers of proteins showing significant (2-fold or higher) fold change in at least one UV-exposed sample are also marked with a star (*).

5.4.4.5 Enzymes involved in repair and replication of DNA

The functional annotation of yeast LEV-2 proteins quantified by MudPIT approach identified 6 enzymes involved in repair and replication of DNA (Figure 5.9). The DNA ligase (Rhomi1_151258) and DNA-directed DNA polymerase (Rhomi1_90583) showed a fold change increase in yeast samples irradiated for 24 hours. The DNA ligase (Rhomi1_151268), DNA-directed DNA polymerase (Rhomi_155605) and DNA helicase (A0A125PJD0) also exhibited a moderate fold change reduction (fold change ~1.5) in yeast samples exposed to 2 hours and 4 hours of UV.

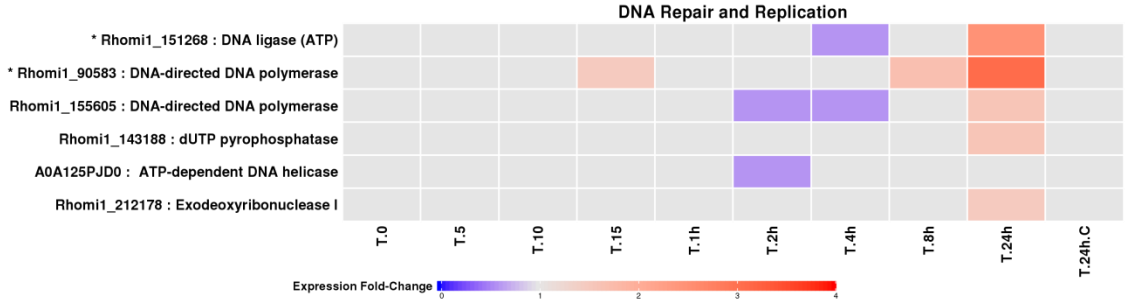


Figure 5.9 Fold change profiles of LEV-2 DNA repair and replication enzymes

Figure displays the heat-map of fold changes of yeast LEV-2 enzymes involved in the repair and replication of DNA. The cultures of yeast LEV-2 were exposed to UV for 5 minutes (T.5) to 24 hours (T.24h), and compared to non-irradiated controls (T.0 and T.24h.C). The proteins showing a fold change reduction in the UV-exposed samples are coloured blue, while proteins showing a fold change increase are coloured red. The proteins exhibiting fold-changes lower than 1.5 are coloured light-grey, and the identifiers of proteins showing significant, 2-fold or higher, fold change in at least one UV-exposed sample are marked with a star (*).

5.4.4.6 Yeast heat-shock proteins and chaperonins

The 24 heat shock proteins (HSPs) were identified and quantified in this study (Figure 5.10). Two small, 10 kDa, HSPs exhibited fold change increase in LEV-2 samples exposed to 8 hours and 24 hours of UV-B. The fold changes of all three identified chaperonin ATPase were increased in LEV-2 cultures exposed to 24 hours of UV-B. The ATPase TCP-1 subunit theta (Rhomi1_150780) also exhibited fold change increase in LEV-2 irradiated for 4 hours and 8 hours, while the ATPase TCP-1 subunits epsilon (Rhomi1_159882) and beta (A0A120E9R4) showed fold change reduction in LEV-2 irradiated for 2 hours. The six 60 kDa HSPs and nine 70 kDa HSPs were quantified; these proteins exhibited a fold change reduction in LEV-2 cultures exposed to 1 hour to 8 hours of UV-B, and some of these proteins (Spoli1_199568, Rhomi1_3145, P19882, A0A0K3CKS4, A0A109FFK7 and P16474) also showed moderate fold change increase in yeasts irradiated for 24 hours. Of the four identified large, 90 kDa, HSPs, all four exhibited high fold change increase in LEV-2 cultures irradiated for 24 hours, while three 90 kDa HSPs (A0A0D6EKA8, A0A109FF04 and Sporo1_21071) also displayed a fold change increase in yeast irradiated for 8 hours, and the HSP A0A109FKA3 also showed a fold change reduction in samples irradiated for 2 hours and 4 hours.

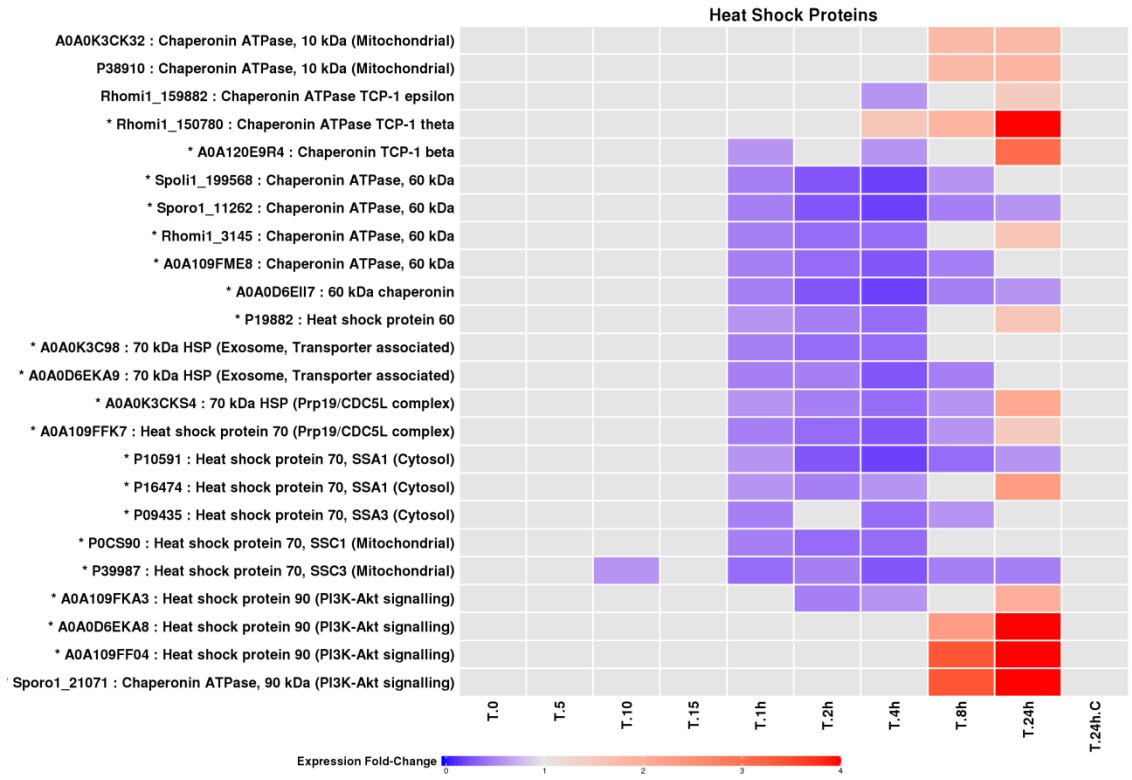


Figure 5.10 Fold changes of LEV-2 heat-shock proteins and chaperonins

Figure displays a heat-map of fold changes of yeast LEV-2 heat-shock proteins and chaperonins. The yeast LEV-2 cultures were exposed to UVR for 5 minutes (T.5) to 24 hours (T.24h), and compared to non-irradiated controls (T.0 and T.24h.C). Proteins with fold change reduction in the UV-exposed cultures are coloured blue, while the proteins exhibiting a fold change increase are coloured red. Proteins with the fold changes lower than 1.5 are coloured light-grey, and the protein identifiers of proteins showing a significant, 2-fold or higher, fold change in UV-exposed samples are marked with a star (*).

5.4.4.7 Proteins involved in biosynthesis of mycosporine-like amino acids

Database matching of mass spectra generated from the proteins isolated from yeast LEV-2 cultures exposed to UV-B radiation and from non-irradiated controls identified 10 proteins involved in the biosynthesis of MAAs. The listed proteins were quantified in all yeast cultures, with the exception of hybrid non-ribosomal peptide synthetase (NRPS) protein, which was not quantified in the LEV-2 cultures exposed to 1 hour and 8 hours of UV (Figure 5.11). Phospho-2-dehydro-3-deoxyheptonate aldolase showed a significant fold change decrease in LEV-2 cultures exposed to UV-B for 1 hour to 24 hours. The hybrid NRPS protein showed a fold change increase in LEV-2 cultures exposed to 5 minutes, 15 minutes, 2 hours and 24 hours of UV-B, and in non-irradiated control grown for 24 hours. The chorismate synthase enzyme showed a fold change increase in the yeast sample exposed to 24 hours of UV-B. All of the five identified transaldolase enzymes exhibited a fold change decrease in LEV-2 irradiated for 2

hours and 4 hours. 3-dehydroquinate synthase exhibited a moderate fold change decrease in yeast culture exposed to 24 hours of UV-B, and a moderate fold change increase in non-irradiated yeast culture grown for 24 hours.

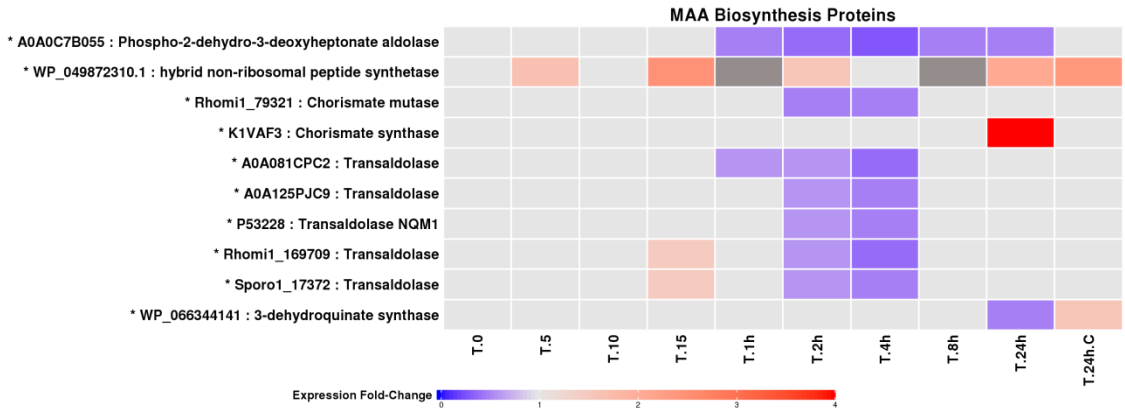


Figure 5.11 Expression profiles of LEV-2 enzymes involved in biosynthesis of MAAs

Heat-map shows the fold changes of yeast enzymes involved in biosynthesis of MAAs. The yeast LEV-2 cultures were exposed to UV for 5 minutes (T.5) to 24 hours (T.24h), and non-irradiated cultures were used as controls (T.0 and T.24h.C). The proteins showing a fold change reduction are coloured blue, while proteins with fold change increase are coloured red. The proteins exhibiting fold change lower than 1.5 are coloured light-grey, and unquantified samples are coloured dark grey. Protein identifiers of proteins showing a significant, 2-fold or higher, fold change are marked with a star (*).

5.5 Discussion

The purpose of this study was to describe, at the proteome level, the stress response mechanisms of a UV-tolerant yeast model, and to determine if the bZip proteins play a role in the stress response of UV-tolerant yeasts. The *Sporobolomyces* yeast LEV-2, previously isolated from the leaves of Brazilian plants (479), was chosen as the model organism for this study based on its high UV-tolerance (Figure 5.1, Figure 5.3), and because it was shown that cell extracts of LEV-2 cultures exhibited an increase in antioxidant, DPPH free radical quenching, activity during long-term exposure to UV-B (Figure 5.2/A). In addition, the previous analysis of this yeast placed it into Division Basidiomycota (479), unlike the commonly studied baker yeast *Saccharomyces cerevisiae*, which belongs to Division Ascomycota. The Ascomycota and Basidiomycota fungi are evolutionary distant, and are considered to have diverged ~650 million years ago (501,502). Thus, this study enabled the comparison of stress response mechanisms between the major, evolutionary distant, divisions of fungi.

The yeast LEV-2 was grown in liquid medium to mid-exponential phase of growth and exposed to UV-B irradiation for up to 24 hours. Solubilised proteins were extracted from the culture after 5 min, 10 min, 15 min, 1 hour, 2 hours, 4 hours, 8 hours and 24 hours of irradiation. A total of 751 proteins could be identified and quantified using TMT-labels and MudPIT high throughput proteomics, and the proteins were functionally annotated by InterProScan (484), KEGG tools (342,482) and literature searches.

5.5.1 UV-B irradiation induces the stress response of yeast LEV-2

In addition to causing direct DNA damage by inducing the formation of DNA photo-adducts (37), UV radiation stimulates the production of reactive, oxygen-derived, chemical species (RS) in the UV-exposed cells, and induces oxidative stress (39,40). The evidence of UV-caused oxidative damage has been found in multiple cell lines and *in-vivo* models. For example, UV-B radiation has been found to induce the production of RS in human cell lines by enhancing the RS generating activity of catalase (39), to induce production of RS in cyanobacteria, as detected by RS-sensing probe 2',7'-dichlorodihydrofluorescein diacetate (503,504), and is considered also to be a major source of oxidative damage to yeast cells (505). UV-A radiation was also found to cause single stranded DNA breaks associated with oxidative damage in mouse models (40) and in human skin (506).

In this study, the functional annotation of 227 proteins showing the fold change increase in yeast LEV-2 cultures exposed to UV-B (Figure 5.3/A) identified a high number of proteins associated with stress-response and cellular signalling (Tables 5.2/A-5.4/A). The increase in fold changes of these proteins in LEV-2 cultures exposed to 8 hours and 24 hours of UV-B (Table 5.3/A) correlated with the increase of antioxidant activity of cell extracts of LEV-2 cultures exposed to 8 hours and 24 hours of UV-B, as measured by the DPPH free radical quenching assay (Figure 5.2/A), suggesting that the yeast LEV-2 increased the rate of biosynthesis of antioxidant enzymes when exposed to UV-B.

While UV-stress responses of UV-tolerant *Sporobolomyces* yeasts have not been previously studied, the stress responses of yeast *Saccharomyces cerevisiae* exposed to various environmental stresses, including heat-shock, pH changes, oxidants such as H_2O_2 and cadmium, hyper-osmotic shock, starvation, and ionizing radiation, have previously been quantified using microarrays (507). These studies identified that the expression levels of large number of mRNA transcripts (~900 genes) change when the yeast is exposed to stress, but the changes in transcript levels are transient, and often adjust to levels close to unstressed cells during the conditions of prolonged stress (507,508). The fold changes of proteins in UV-B

exposed yeast LEV-2 cultures (Figure 5.4) match the patterns observed in these microarray studies of yeasts exposed to environmental stresses (507,508), suggesting that the stress response mechanisms are conserved between yeast *S. cerevisiae* and *Sporobolomyces* yeast LEV-2.

The conservation of stress response patterns between yeasts of Genus *Saccharomyces* (Division Ascomycota) and Genus *Sporobolomyces* (Division Basidiomycota) would indicate that the stress response mechanisms are conserved between fungal lineages that diverged approximately 650 million years ago (501,502), and that the majority of Basidiomycota and Ascomycota fungi possess homologous stress response mechanisms. This is further supported by the recent proteomics study by Villegas et al (2014), which examined the stress response of yeast of the Genus *Rhodotorula* (Division Basidiomycota) exposed to oxidative stress induced by high levels of copper, and identified an increase in expression levels of stress response proteins such as heat shock proteins and superoxide dismutase in stressed yeasts (509). This similarity of stress response in evolutionary distant groups of fungi is possibly because the stress signalling mechanisms evolved during the early evolution of eukaryotic life. Our previous phylogenetic studies of evolution of the Nrf2 pathway suggested that the bZip-protein based response to oxidative stress evolved in early eukaryotes as an adaptation to oxidative environment caused by the rising oxygen levels during the geological time (510). The expression fold change patterns of stress response proteins, including the bZip proteins and other signalling proteins of UV-stressed *Sporobolomyces* yeast LEV-2 were further examined to evaluate if the stress response is conserved between the yeast *S. cerevisiae* and LEV-2, and the results are discussed in the following section.

5.5.2 Stress response proteins of yeast LEV-2

MudPit analysis of UV-irradiated cultures of *Sporobolomyces* yeast LEV-2 quantified 105 proteins involved in cellular stress response (section 5.4.4). Stress response proteins were functionally annotated using the InterProScan, KEGG tools and manual annotation.

5.5.2.1 Basic-leucine zipper proteins

In animals, the bZip transcription factor Nrf2 activates the transcription of a large number of genes (125,129) encoding detoxification and antioxidant enzymes such as aldehyde dehydrogenases, glutaredoxins and thioredoxins, as well as enzymes involved in biosynthesis of glutathione (96). The function of the bZip transcription factor Nrf2 was characterized by high-throughput technology such as Chip-seq and microarrays (125,129), and its role in protection against oxidative stress was confirmed in cell lines and animal models (96). For

example, genetically modified mouse models deficient in Nrf2-encoding gene were found to be highly sensitive to carcinogens such as benzo[a]pyrene (184), to environmental pollutants such as diesel exhaust (183), and to toxicity of drugs such as acetaminophen (185). In addition, pharmaceuticals that activate Nrf2-signalling, such as sulforaphane (SFN) and butylated hydroxyanisole, were shown to activate the transcription of Nrf2-regulated genes, such as *GST*, *GCLC* and *NQ01,* in wild type mouse models, but not in Nrf2-knockout models (122). The activation of Nrf2 by SFN has also been shown to protect the human cell models and mouse animal models against UV-induced oxidative stress (467,470,511). The bZip proteins homologous to vertebrate Nrf2 protein have also been studied in invertebrates such as the fly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*, and were found also to regulate the response to oxidative stress in these invertebrate models (325,329,512).

The genome of baker yeast *S. cerevisiae* encodes several bZip transcription factors including Gcn4 and eight Yap proteins (Yap1 – Yap8). The bZip protein Gcn4 activates the biosynthesis of amino acids, and is involved in response to starvation, but also in UV-stress (472), as evinced from low UV-tolerance of Gcn4 knockout yeast models (471). The bZip transcription factor Yap1 is a major regulator of oxidative stress response in *S. cerevisiae,* while Yap2, Yap5 and Yap8 proteins play a role in response to metal-induced stress, Yap4 and Yap6 in regulating response to osmotic stress, and the roles of Yap3 and Yap7 are currently unknown (475). The function of Yap1 was inferred because *YAP1*-knockout yeasts, but not yeasts deficient in other *YAP* genes, were found to possess low activity of antioxidant enzymes such as superoxide dismutase and glutathione reductase, and display low adaptability to oxidative stress (513). In addition, the Yap1-binding motif on DNA has been found in promotor regions of antioxidant genes such as *GSH1* and *TRX2* (475). The homologs of Yap1 protein have been empirically validated in yeasts *Candida albicans* (Cap1 protein) and *Schizosaccharomyces pombe* (Pap1 protein) (514), suggesting that bZip-regulated response to oxidative stress is conserved between yeasts of Division Ascomycota. While the Yap1 homologs have not been experimentally confirmed in *Sporobolomyces* (Division Basidiomycota) yeasts, our previous bioinformatics study found that genomes of Basidiomycota fungi encode homologs of bZip proteins similar to animal bZip transcription factor Nrf2 (408), and the recent study by Jindrich and Degnan (2016), identified also that animal bZip-encoding genes evolved from the bZip-encoding gene of unicellular eukaryote ancestor of fungi and animals (515).

MudPIT proteomics analysis of *Sporobolomyces* yeast LEV-2 identified four bZip proteins similar to fungal bZip transcription factors (Figure 5.5), suggesting that bZip proteins are conserved between *S. cerevisiae* and *Sporobolomyces* yeasts. The LEV-2 bZip protein,

designated LEV-2_XP_748177.1, showed a significant fold change increase in UV-irradiated LEV-2 cultures (Figure 5.5), suggesting that LEV-2_XP_748177.1 is a LEV-2 homolog of *S. cerevisiae* Yap1 and plays a role in stress signalling in the yeast LEV-2. This is because the expression of Yap1 is increased in *S. cerevisiae* cultures exposed to oxidants such as H_2O_2 and paraquat (516), while the expression of other *S. cerevisiae* Yap proteins is not changed significantly during oxidative stress (475,517). The fold change patterns of bZip protein LEV-2_XP_748177.1 in yeast LEV-2 exposed to long term UV-B irradiation (Figure 5.5) were followed by the fold change increase in stress-response proteins of yeast LEV-2 (Figure 5.4), and these fold changes showed striking resemblance to changes in expression levels of bZip transcription factor Nrf2 in human liver cancer (HepG2) cells and rat renal epithelial cells exposed to electrophiles such as tBHQ and β-NF (518) or to heme (519). In these studies, exposure to Nrf2 activator caused an increase in cellular concentration of Nrf2 after ~30 minutes to 1 hour, followed by an increase in expression of Nrf2-regulated genes after 2 or more hours.

While the observed protein fold changes in LEV-2 cultures exposed to UV-B suggest that the bZip protein LEV-2_XP_748177 of *Sporobolomyces* yeast LEV-2 is a homolog of *S. cerevisiae* bZip transcription factor Yap1 and of vertebrate bZip transcription factor Nrf2, other signalling proteins might also be involved in stress response in yeast LEV-2. For example, RAS/cAMP/PKA signalling and bZip protein Gcn4 are known to be involved in UV-response of yeast *S. cerevisiae* (471). Thus further studies, potentially involving gene deletion of LEV-2_XP_748177-encoding gene in yeast LEV-2 or gene-knockdown by siRNA, are required to exclude the possibility that the observed changes in proteome of yeast LEV-2 exposed to UV-B radiation are mediated by other signalling pathways. It should be noted also that the genome of yeast LEV-2 is currently not available, and the presence of bZip proteins in LEV-2 was inferred by database matching of tandem mass spectra generated from tryptic digests of LEV-2 proteins. Therefore, primary amino-acid sequence of discovered bZip proteins could not be established, and genome sequencing yeast LEV-2 is required to elucidate the DNA sequence of bZip protein encoding genes of yeast LEV-2. The assembly of LEV-2 genome would facilitate the phylogenetic analysis of LEV-2 bZip proteins and sequence similarity comparisons of LEV-2 bZip proteins to vertebrate and known yeast bZip proteins. In addition, the knowledge of DNA sequence of bZip-encoding genes of the yeast LEV-2 would allow the design of PCR primers for real-time PCR analysis of bZip-encoding mRNA transcripts in yeast LEV-2 exposed to oxidative stress, and the design of siRNAs for knock-down experiments.

In addition to the XP_748177.1 like protein, three other bZip proteins were identified and quantified in the proteome of yeast LEV-2, and the expression levels of these proteins were reduced or unchanged in UV-B irradiated yeast cultures (Figure 5.5). These proteins are likely to be homologs of yeast *S. cerevisiae* Yap proteins (Yap2 – Yap8) not involved in response to oxidative stress. This is because the yeast *S. cerevisiae* Yap2 – Yap8 proteins do not show significant fold changes in *S. cerevisiae* exposed to oxidants (475,507,508,517), and do not regulate the response to oxidative stress (473). For example, Yap5 protein is involved in iron metabolism (520), Yap6 regulates the response to osmotic stress, and functions of Yap3 and Yap7 are currently unknown (475).

5.5.2.2 Signalling and apoptosis related proteins

The proteomics study of LEV-2 yeast identified 22 proteins associated with different cellular signalling pathways. These proteins were further classified based on the pathway:

The 14-3-3 proteins

The 14-3-3 proteins are involved in numerous cellular processes, including signal transduction, cell-cycle control and apoptosis, and signalling roles of these proteins are an active field of research (521). Three 14-3-3 proteins were identified in this study, and the fold changes of these proteins were significantly reduced by moderately long UV-B irradiation (1 hour to 4 hours of UV-B exposure). This reduction in expression co-occurs with increase in cell death observed in samples under 1 hour to 4 hours of UV-B (Figure 5.1), suggesting the link between levels of 14-3-3 proteins and cell death, possibly by UV-induced apoptosis. This is in agreement with studies of Zhang et al. (1999), who identified a strong correlation between expression of 14-3-3 proteins and apoptosis in HeLa cell line model (522). It is currently unknown whether 14-3-3 proteins play identical roles in human cell line models and in yeasts, and different stresses were reported to affect the expression of 14-3-3 proteins in yeast *S. cerevisiae* in different fashion, depending on the source of stress. For example, Yoshimoto et al. (2002) found that calcium induced stress caused the reduction in expression levels of 14-3-3 proteins in *S. cerevisiae*, as measured by microarrays (523). The similar microarrays-based study by Gasch et al. (2002) identified that changes in levels of 14-3-3 proteins depend on the source of stress, with protein levels increased in yeast *S. cerevisiae* exposed to heat-shock, and reduced in the sample exposed to H_2O_2 (508).

MAPK signalling proteins

The mitogen-activated protein kinase (MAPK) signalling pathway mediates transduction of extracellular signals, and is essential for multiple cellular functions, such as cell differentiation, proliferation, initiation of apoptosis and adaptation to environmental stresses (491). The basic assembly of MAPK signalling comprises three kinases - MAPK kinase kinase (MAPKKK), MAPK kinase (MAPKK), and MAP kinase (MAPK) - that transduce extracellular signal by sequential phosphorylation reactions. MAPKKK phosphorylates MAPKK, which in turn phosphorylates MAPK, and the phosphorylated MAPK phosphorylates transcription factors such as Sko1 which control transcription of genes involved in stress response (524). This three-component module is conserved between yeasts and animals (524), and the increase in expression of MAPK kinases has been associated with increase in resistance to oxidative stress in yeasts *S. cerevisiae* and *S. pombe* (514), invertebrates such as *C. elegans* (525), mouse animal models and human cell lines (524,526). For example, *S. cerevisiae* and *S. pombe* yeasts deficient in MAPK kinase encoding genes are hypersensitive to salt-induced stress, heat shock and nutritional limitation (514); in *C. elegans*, oxidative stress induced by sodium arsenite, paraquat, or t-butyl peroxide increases the expression of MAPK kinase p38, while the deletion of gene encoding the MAPK kinase SEK-1 increases sensitivity to oxidative stress (525); human HeLa cells exposed to H_2O_2-induced oxidative stress exhibit an increase in expression of multiple MAPK kinases, and the inhibition of extracellular MAPK signalling kinases increases the apoptosis rate in stressed cells (526).

The cell division control proteins 42 (Cdc42) were shown to be involved in activation of yeast high osmolarity glycerol (HOG) MAPK pathway (527). The HOG pathway involves a series of kinases which activate Hog1 MAP kinase in response to severe osmotic stress, and the mutation of yeast *S. cerevisiae* gene encoding Cdc42 proteins has been shown to inhibit the yeast response to osmotic stress, while mutants over-expressing Cdc42 proteins have been found to exhibit an increase in stress response (527). This is because Cdc42 proteins interact with, and activate, the Ste20 MAP4K component of HOG pathway in yeast, and this interaction is essential for signal transduction during osmotic stress (528). This function of Cdc42 proteins in MAPK pathway is conserved between yeast and mammals, and it has been shown that stress-activated MAPK pathway kinases such as JNK1 and p38 are activated by the Cdc42 protein in human tissue extracts (529) as well as monkey kidney cell (COS-1) cultures (530).

Four kinases associated with MAPK signalling pathways, and two Cdc42 proteins, were identified by MudPIT analysis of *Sporobolomyces* yeast LEV-2. These six proteins showed fold change increases in yeast cultures exposed to long term, 8 hours or longer, UV-B radiation

(Figure 5.6). These results suggest that a MAPK signalling is activated by UV-B radiation in yeast LEV-2, and that these signalling pathways are evolutionary conserved between Ascomycota yeasts such as *S. cerevisiae* and *S. pombe*, and the *Sporobolomyces* yeast LEV-2. While MAPK pathway has not yet been studied in *Sporobolomyces* yeasts, this pathway was empirically characterized and found to mediate a response to stress in Basidiomycota fungi *Cryptococcus neoformans* exposed to fungicide fludioxonil and to high levels of NaCl (531). In this study, the increase in MAPK signalling kinases was found to correlate with the fold change increase of enzymatic antioxidants, such as SODs (Figure 5.8), and with the fold change increases of certain heat shock proteins (Figure 5.10) observed in LEV-2 cultures exposed to 8 hours and 24 hours of UV-B radiation. In addition, the cell extracts of LEV-2 cultures irradiated for 8 hours and 24 hours also showed the increase in DPPH free radical scavenging activity (Figure 5.2). These results, and the assumption that MAPK-mediated stress response is conserved across all eukaryotes (524) indicate that MAPK signalling is likely involved in activating stress response in UV-irradiated yeast LEV-2. It is however, known that other signalling pathways, such as Yap1 signalling (475), are also involved in yeast stress response, and further empirical studies, such as inactivation of genes encoding MAPK kinases in yeast LEV-2, are required to establish the importance of MAPK signalling in yeast LEV-2 in comparison to other signalling pathways, such as the Yap1 pathway.

FoxO signalling

FoxO signalling proteins represent a subfamily of the forkhead family of transcription factors, and are highly conserved across all animal phyla, with orthologs discovered in cnidarian *Hydra vulgaris*, worm *C. elegans*, fly *D. melanogaster*, mouse and rat models and human cell lines (532). These proteins have been associated with resistance to oxidative stress, control of life-span, regulation of cell cycle arrest, and induction of apoptosis (532). In the nematode *C. elegans*, FoxO signalling is mediated by the forkhead transcription factor Daf-16, and the increase in expression of this protein was shown to increase the worm life-span and resistance to oxidative stress (533). The increase in FoxO signalling was found also to mediate stress response in a mammalian cell line: H_2O_2-induced oxidative stress caused the fold change increase of the p66shc protein, mammalian homolog of Daf-16, while cells deficient in p66shc-encoding gene were found to be highly sensitive to H_2O_2 (534).

The yeast *S. cerevisiae* forkhead proteins HCM1, FKH1 and FKH2 proteins are homologs of animal FoxO proteins, and have been shown to regulate stress response and longevity of *S. cerevisiae* (490,535,536). For example, the study by Postnikoff et al. (2012) found that *S. cerevisiae* mutants deficient in genes encoding FKH1 and FKH2 proteins had a shorter life-

span and lower resistance to H_2O_2 than wild-type yeasts, while *S. cerevisiae* genetically modified to overexpress FKH proteins showed an increase in life-span and resistance to H_2O_2 (490). The study by Maoz et al. (2015) also found that *S. cerevisiae* genetically modified to overexpress FoxO homolog HCM1 had a high rate of transcription of genes encoding catalase and superoxide dismutase enzymes; in addition, these yeasts were highly resistant to oxidative stress induced by H_2O_2, and had a higher lifespan than wild-type yeasts (535).

Two FKH proteins were identified and quantified in this study (Figure 5.6), and levels of these proteins were moderately increased (fold change ~1.5) in the yeast LEV-2 cultures exposed to long-term UV-B radiation (8 hours and 24 hours). This increase in expression of FKH proteins co-occurred with the increase of free radical quenching activity of cell extracts of LEV-2 yeast cultures exposed to UV-B radiation (Figure 5.2) and with a significant fold change increase of superoxide dismutase enzymes (Figure 5.8). These expression patterns match the results of previous studies that quantified the expression of FKH proteins of yeast *S. cerevisiae* exposed to H_2O_2-induced stress (490,535), and imply that a fold change increase in FKH proteins is associated with yeast LEV-2 resistance to UV-induced oxidative stress. These results also suggest that the FoxO pathway is conserved between yeast *S. cerevisiae* and *Sporobolomyces* yeasts such as LEV-2.

Ras signalling proteins

In yeasts, such as the model yeast *Saccharomyces* cerevisiae, the Ras signalling pathway controls the DNA damage-independent response to UV-induced stress. This stress response pathway is comprised of membrane associated Ras proteins that activate the adenylate cyclase enzymes to stimulate the production of cyclic AMP (cAMP). The increased cytosolic concentration of cAMP activates a protein kinase A (PKA) controlled phosphorylation cascade that increases the translation of bZip transcription factor Gcn4 (471), and leads to induction of genes involved in biosynthesis of amino acids (472). The yeast Ras signalling pathway is considered to be homologous to mammalian UV-response pathway that includes Ras associated proteins Ha-Ras and Raf-1 as well as transcription factors such as NF-kB and AP-1 (471,537), and is distinct from yeast response to DNA damage, which is mediated by the DNA damage responsive protein kinase Dun1 (477). This is because Ras signalling involves membrane associated Ras proteins rather than DNA damage sensing kinases. In addition, Ras signalling is involved in regulation of amino acid biosynthesis, but not in regulation of DNA repair, and transcription of Ras associated proteins is induced by UV in *S. cerevisiae* strains deficient in Dun1-encoding gene (471). While primarily associated with response to starvation and with induction of amino acid biosynthesis (538), Ras signalling was shown to also regulate the

protection against UV-induced stress in the yeast *S. cerevisiae*. This is because the *S. cerevisiae* strains deficient in gene encoding transcription factor Gcn4 are highly sensitive to UV irradiation, while the strains engineered to constitutively over-express Gcn4 are highly resistant to UV (471).

In this study, the Ras-associated proteins Rap-7A and Rap-11A of *Sporobolomyces* yeast LEV-2 showed a moderate, 1.5-fold, increase in LEV-2 yeast cultures exposed to 24 hours of UV-B (Figure 5.6), suggesting that Ras signalling is enhanced in yeast LEV-2 during long-term UV-B irradiation. This increase in fold change co-occurred with the low death rate observed for LEV-2 cultures exposed to 8 hours and 24 hours of UV-B (Figure 5.3), suggesting that the Ras-associated proteins of yeast LEV-2 induce the adaptation to UV-B stress, and that the function of Ras-associated proteins is conserved between *S. cerevisiae* and the yeast LEV-2. The observed reduction in yeast LEV-2 cell death is, however, also associated with an increase in expression of enzymatic antioxidants (Figure 5.8) and other signalling pathways involving bZip transcription factors (Figure 5.5), FoxO signalling, and MAPK signalling (Figure 5.6). Thus, further research is required to quantify the contributions of individual signalling pathways to the stress response of the UV-exposed yeast LEV-2.

The Ras-related calcium-binding protein calmodulin is an important component of stress signalling in the yeast *S. cerevisiae*, and yeast strains deficient in the calmodulin-encoding gene have been shown to be sensitive to stress induced by elevated levels of ions such as OH^-, Mn^{2+}, Na^+ and Li^+ (539). The calmodulin and Ca^{2+}/calmodulin-dependent phosphatase calcineurin mediate the response to elevated levels of cytosolic Ca^{2+} caused by oxidative stress, osmotic shock and other sources of stress (539,540). Calcineurin responds to stress-associated elevation of Ca^{2+} levels in the cell by dephosphorylating the transcription factor Crz1p. The dephosporilated Crz1p translocates to the nucleus, where it activates transcription of genes encoding numerous proteins involved in cell survival (539). For example, the microarray study of Yoshimoto et al. (2002) identified that over 150 genes are under transcriptional control by Crz1p in the yeast *S. cerevisiae*, including genes involved in control of ion transport and homeostasis, cell wall synthesis/maintenance, lipid and sterol metabolism, vesicle transport, and cellular signalling (523). The stress response function of calmodulin signalling is conserved between the baker yeast *S. cerevisiae* and the pathogenic yeast *Candida albicans*, and it was shown that *C. albicans* deficient in genes encoding calmodulin-dependent protein kinases is highly sensitive to H_2O_2-induced oxidative stress (541). This pathway, however, is not involved in the stress response of fungus *Aspergillus nidulans* (540), indicating that the function of calmodulin is not conserved across the different fungal phyla. In this study, the LEV-2

calmodulin protein (Figure 5.6) showed a fold change increase in yeast cultures exposed to 1 hour to 4 hours of UV-B, and showed the pattern similar to LEV-2 bZip transcription factor, which also showed a fold change increase in UV-B exposed LEV-2 cultures (Figure 5.5). Studies of different cell models such as hepatocytes, lymphocytes and endothelial cells, showed that oxidative stress disrupts the function of cellular Ca^{2+} transporters and increases the concentration of Ca^{2+} in the cytosol (542). Thus, the observed fold change increase in the yeast LEV-2 calmodulin protein suggests that the calmodulin activity is increased in the yeast LEV-2 in response to an increase in cytosolic concentration of Ca^{2+} due to UV-induced oxidative stress. This would indicate that calmodulin is an early sensor of cellular stress in yeast LEV-2, and the function of this protein is conserved between yeast *S. cerevisiae* and *Sporobolomyces* yeast LEV-2.

Cell cycle control and apoptosis

Yeasts, and other unicellular organisms, were traditionally considered not to possess the mechanisms for induction of programmed, apoptotic, cell death, but the expression of mammalian proapoptotic genes such as *BAX* and *TP53* was shown to induce the apoptotic cell death in yeast *Saccharomyces cerevisiae* (543,544). In addition, it was shown that a moderate concentration of H_2O_2 (3 – 5 mM) induces apoptotic cell death in *S. cerevisiae*, while a high concentration of H_2O_2 (180 mM) causes necrotic cell death (543). The discovery of the *S. cerevisiae* caspase *YCA1* and the yeast apoptosis-inducing factor-1 (*AIF1*), and studies that demonstrated that cell death is reduced in *S. cerevisiae* yeasts deficient in genes encoding these proteins, further confirmed that certain apoptotic cell death mechanisms are conserved between yeasts and animals (545). The yeast LEV-2 homologs of genes encoding Yca1 and Aif1 proteins were not discovered in this study, possibly because amino acid sequences of these proteins in yeast LEV-2 are significantly different from the proteins of fungi for which proteomes were available, and which were used for database matching of tandem mass spectra generated from the tryptic fragments of LEV-2 proteins. The cell division proteins 48 and Ras associated proteins were, however, successfully identified and quantified (Figure 5.6), and these proteins are known to be involved in apoptosis in yeast *S. cerevisiae* (546,547).

The cell division proteins 48, referred to as Cdc48 in yeasts, and p97 proteins in animals, are multipurpose proteins conserved across fungi and animals, and essential for growth of the yeast *S. cerevisae* (548). These proteins have been associated with numerous functions including protein degradation, protein aggregation, control of cell cycle and apoptosis, and transcription and replication of DNA (548). The yeast Cdc48 proteins, and animal homologs, were also shown to play a role in endoplasmic reticulum (ER) stress; these proteins are involved in

extraction of misfolded proteins from the ER into the cytosol, where the misfolded proteins are subsequently degraded (549). Mutation in the *S. cerevisiae CDC48* gene has been shown to increase the rate of apoptosis in yeast (547), and an increase in apoptosis was also shown in human cell line, and in the fly *Drosophila melanogaster* when the translation of VCP protein, the animal homolog of Cdc48, is knocked-down by siRNAs (550). The fold change reduction of Cdc48 proteins of yeast LEV-2 was observed in yeast cultures exposed to 2 hours and 4 hours of UVR (Figure 5.6), coinciding with the fold change decrease in proteins associated with catabolism of misfolded proteins, protein folding and protein refolding (Table 5.2/A, Figure 5.4/B). In addition, the cell death rate of yeast LEV-2 cultures exposed to 2 hours and 4 hours of UV-B was considerably higher than the death rate of LEV-2 cultures exposed to 8 hours or 24 hours of UV-B (the time period where the Cdc48 proteins do not show fold change decrease). These results suggest that Cdc48 proteins are linked to misfolded protein stress, and possibly to apoptosis, in *Sporobolomyces* yeast LEV-2, and imply that function of Cdc48 proteins is conserved between *Basidiomycota* fungi and *Ascomycota* fungi.

In addition to the UV response, Ras and Ras-related (Rap) proteins have been associated with control of cell cycle and apoptosis in yeast *S. cerevisiae* (551), and constitutive high expression of Ras signalling proteins was demonstrated to reduce the lifespan, and induce the apoptotic cell death, of yeast *S. cerevisiae* (552,553). The study by Gourlay and Ayscough (2006) found that constitutive activation of Ras signalling, induced by knock-out of genes that encode the Ras regulating proteins Sla1p or End3p, leads to an increase in cAMP levels in yeast, followed by the accumulation of RS in the cytosol and apoptotic cell death (553). Similar results were reported by Heeren et al. (2004), who found that mutations in *RAS* genes of *S. cerevisiae* led to redox imbalance characterized by the excretion of cytosolic glutathione, elevated RS concentration, reduction in yeast life-span, and apoptotic cell death (552). The activation of apoptosis by Ras/cAMP/PKA signalling was also demonstrated for the pathogenic yeast *Candida albicans*, where the mutations that block Ras signalling, such as deletions of *RAS1* and *CDC35* genes, were found to suppress the apoptotic response to acetic acid or H_2O_2, while mutations that stimulate Ras signalling, such as $RAS1^{val13}$ mutation, accelerated the rate of apoptosis when *C. albicans* was exposed to acetic acid or H_2O_2 (554) . Unlike yeasts, the constitutive expression of Ras signalling has been associated with carcinogenesis in animals. For example, mutations in *HRAS*, *KRAS* and *NRAS* oncogenes that lead to constitutive transcription of Ras-controlled genes have been associated with human cancers and carcinogenesis in animal models (555).

The MudPIT study of the UV-B irradiated *Sporobolomyces* yeast LEV-2 quantified six Ras-associated proteins (Figure 5.6). The levels of these proteins, however, did not show significant fold change increase in LEV-2 exposed to 1 hour - 4 hours of UV-B irradiation, which is the time period of high cell death rate (Figure 5.3). This result indicates that UV-B induced cell death of yeast LEV-2 is not due to over-expression of Ras signalling, and is likely caused by other mechanisms, such as UV-induced generation of RS leading to oxidative stress. In addition, the yeast LEV-2 death rate (Figure 5.3) showed a marked decrease in yeast LEV-2 cultures exposed to 8 hours of UV-B and 24 hours of UV-B; this reduction in cell death coincided with the fold change increases in enzymatic antioxidants (Figure 5.8), 10 kDa and 90 kDa heat shock proteins (Figure 5.10), and proteins involved in the MAPK, Ras and FoxO signalling pathways (Figure 5.6). Furthermore, the antioxidant activity of cell extracts of LEV-2 yeasts, as measured by DPPH free radical quenching was also highly elevated after 8 hours and 24 hours of UV-B exposure (Figure 5.2). These results indicate that the increase in antioxidant activity is linked with reduction in yeast LEV-2 cell death, suggesting that the primary cause of cell death of yeast LEV-2 exposed to UV-B is oxidative stress, possibly leading to apoptosis, rather than direct, UV-inflicted, DNA damage. While this is consistent with previous studies that identified that moderate oxidative stress induces apoptosis rather than necrosis (543) in yeast *S. cerevisiae*, it should be noted that measurement of yeast viability after UV-B irradiation is not sufficient to differentiate between apoptotic cell death and necrotic cell death, and a recent study of yeast apoptosis recommend a combination of assays to determine the rate of yeast apoptosis (545). Thus, further testing, preferably by a combination of assays to determine the cell viability, accumulation of RS, DNA fragmentation, and cell integrity, are required to measure the ratio of necrotic to apoptotic death in LEV-2 yeasts exposed to UV-B.

Other signalling proteins

Five protein kinases, not related to MAPK pathway, were identified in the yeast sample (Figure 5.6). Of those, three adenylate kinases had low expression levels in samples subjected to 1 hour or longer UV irradiation, while two kinases of unknown specificity, exhibited significantly increased (fold change > 2) expression levels in yeast sample exposed to 24 hours of UVR. The protein kinases are involved in a large number of signalling pathways (556), and further research is required to elucidate the functions of these proteins in LEV-2.

5.5.2.3 Proteins involved in biosynthesis of antioxidants

MudPIT analysis of yeast LEV-2 identified and quantified 17 proteins involved in biosynthesis of small-molecule antioxidants such as glutathione and ubiquinol. The identified proteins (Figure 5.7) included enzymes involved in biosynthesis of glutathione (492) such as hydroxyacylglutathione hydrolases (HAGHs) and glutathione S-transferases (GSTs), enzymes involved in biosynthesis of vitamin B6 (PdxS/SNZ family lyases and pyridoxine 4-dehydrogenases) (493), and succinate dehydrogenase (SdHs) enzymes, which reduce oxidised coenzyme Q10 to its reduced form (ubiquinol) (494).

UV-mediated induction of oxidative stress and depletion of cellular small molecule antioxidants was previously observed in multiple models including animal cell lines, plants, yeasts and bacteria (39,40). For example, Heck et al. (2003) found that UV-B light is absorbed by catalase and increases the H_2O_2 generating activity of this enzyme in human and mouse keratinocytes, and in hamster fibroblasts (39). A study by Podda et al. (1998) found that simulated solar UV radiation depletes small molecule antioxidants ubiquinol, α-tocopherol and ascorbic acid in the cell culture of human skin, in a linear, dose-dependent fashion (557). UV radiation was shown also to induce RS generation in cyanobacteria (558), algae (559), and plants (560), and is also considered to be a major source of oxidative stress in yeasts (505).

In this study, the fold changes of the majority of enzymes involved in biosynthesis of small molecule antioxidants were moderately reduced in the LEV-2 yeast cultures irradiated for 1 hour to 8 hours (Figure 5.7). This fold change reduction co-occurred with the reduction in antioxidant activity of cell extracts of LEV-2 yeast cultures exposed to 1 to 4 hours of UV-B, as measured by DPPH free radical quenching assay (Figure 5.2), indicating that the UV-B radiation, and the associated oxidative stress, deplete the small molecule antioxidants of *Sporobolomyces* yeast LEV-2. These results are in agreement with the reports of Gasch et al. (507,508), who used microarrays to study the protein expression of *S. cerevisiae* subjected to environmental stresses such as heat shock and exposure to H_2O_2 and found that temperature-induced stresses induced a fold change reduction in GST, SdHs and SNZ proteins (507,508) of *S. cerevisiae*, consistent with the fold change patterns observed in this study. While the studies by Gasch et al. (507,508) did not measure the effects of UV-B radiation, it has been shown previously that heat-shock, similarly to UV-B, induces oxidative stress in yeasts. That is because *S. cerevisiae* strains deficient in genes encoding antioxidant enzymes, such as catalase, SoD and cytochrome c peroxidase, are highly sensitive to heat-shock, while yeasts engineered to over-express these enzymes have an elevated resistance to heat shock and oxidative stress (561).

Notably, certain enzymes involved in biosynthesis of small-molecule antioxidants, such as glutamate cysteine ligase (GCL) and GSH synthetase (GS) (562), were not identified in this study. This is possibly due to the lack of the genome sequence for LEV-2 sample, which necessitated the use of related yeast proteomes for the database matching of MS spectra generated from LEV-2 proteins, and likely reduced the number of identified proteins.

5.5.2.4 Enzymatic antioxidants

Enzymatic antioxidants include enzymes that convert RS into less reactive chemical species, such as superoxide dismutase (SOD), superoxide reductase (SOR) and catalase (CAT); enzymes involved in recycling of non-enzymatic small molecule antioxidants (e.g. glutathione reductase); and enzymatic systems that reduce oxidised cellular macromolecules, such as thioredoxin and glutaredoxin systems (54,68). While the individual enzymatic antioxidants are non-essential, presumably because cellular antioxidant systems have a considerable level of redundancy, the yeast *S. cerevisiae* (563,564) and mouse animal models (11) deficient in certain antioxidant enzymes, such as mitochondrial SODs, were shown to be hypersensitive to oxidative stress. In addition, animal cell lines in which the biosynthesis of enzymatic antioxidants was induced by Nrf2 activators such as sulforaphane were found to be highly resistant to UV-induced oxidative stress (468) and to oxidants such as H_2O_2 (565). Analogous increase in resistance to oxidative stress was also found in mouse models exposed to the Nrf2 activator sulforaphane (469), and in yeasts genetically engineered to over-express the bZip transcription factor Yap1 (566). Enzymatic antioxidants are highly conserved across the domains of life, and the expression of human superoxide dismutase enzyme in yeast found was found to increase yeast resistance to oxidants such as paraquat (567).

Aldehyde dehydrogenases (ADHs) are involved in the metabolism of toxic aldehydes produced during oxidative stress (497). One ADH and a benzaldehyde dehydrogenase were identified and quantified by the MudPIT analysis of *Sporobolomyces* yeast LEV-2, and both enzymes showed fold change increase in LEV-2 cultures exposed to 24 hours of UV-B irradiation (Figure 5.8). Superoxide dismutases (SODs) catalyse the conversion of a superoxide anion to H_2O_2 and O_2, and play a critical role in protection against oxidative stress in all eukaryotes, including yeasts (496). Four SODs were identified in this study, and the expression levels of SODs were increased in UV-B irradiated LEV-2 cultures (Figure 5.8). This fold change increase of ADHs and SODs correlated with the increase in antioxidant, DPPH quenching, activity of extracts of yeast LEV-2 cultures exposed to 8 hours and 24 hours of UV-B (Figure 5.2/A), and with the observed reduction in yeast LEV-2 cell death (Figure 5.3), indicating that these enzymes protect yeast LEV-2 against UV-induced oxidative stress. These results are consistent with

previous studies of yeast *S. cerevisiae* exposed to oxidants such as H_2O_2 and to heat-shock induced oxidative stress (508), and suggests that the function of SODs and ADHs is conserved between *S. cerevisiae* and the yeast LEV-2.

Catalases (Cat) facilitate the breakdown of H_2O_2 to O_2 and H_2O, and are major antioxidant enzymes in yeasts (496). A catalase identified in this study showed a moderate fold change reduction in yeast cultures irradiated for 2 hours to 8 hours ((Figure 5.8). Cytochrome C peroxidases (CCPs) catalyse the conversion of H_2O_2 to H_2O, and have been implicated in yeast response to heat shock and oxidative stress (496,498). All of the five identified CCPs had reduced expression levels in yeast samples exposed to moderate duration of UV-B doses (1 hour to 4 hours), and the results were ambiguous for LEV-2 cultures irradaated for 8 hours and 24 hours. Isocitrate dehydrogenase (IDH) enzymes are involved in citric acid cycle, and catalyse the two-step oxidative carboxylation of isocitrate to α-ketoglutarate. IDHs exist in multiple isoforms which differ in the use of NAD+ or NADP+ as a cofactor, and in cellular localization to cytosol or mitochondria. While not direct antioxidants, IDHs play a part in cellular defences against RS by reducing NAD(P)+ to NAD(P)H to enable the regeneration of glutathione (GHS) and thioredoxins (495,500). This study identified two NAD+ dependant IDHs, both of which showed fold change increase in LEV-2 cultures irradiated for 8 hours and 24 hours, and four NADP+ associated IDHs which displayed fold change reduction after 1 hour or longer UV-B exposure. The observed fold change reduction of LEV-2 IDH, CCP and Cat enzymes corresponds to the reduction in DPPH quenching observed for LEV-2 cultures exposed to 1 hour to 4 hours of UV-B (Figure 5.2), and is possibly a result of depletion of these enzymes by UV-B induced oxidative stress. Oxidative stress was reported to deplete antioxidants, especially GSH, in the yeast *S. cerevisiae* (568), but the extent of antioxidant depletion varied with the source and duration of stress (496,508). The depletion of GSH was reported also for UV-irradiated human skin models (557), indicating that oxidative stress has a similar effect on enzymatic antioxidants in diverse eukaryotic organisms.

Glutathione peroxidase (Gpx) enzymes catalyse conversion of RS, such as H_2O_2, to non-reactive compounds such as H_2O, and are major eukaryotic enzymatic antioxidants (499). Two Gpx enzymes were identified in this study, and the lack of major fold changes of these enzymes (Figure 5.8) indicates that Gpx enzymes are stable during the oxidative stress. This is possibly because these enzymes catalyse RS conversion ($ROOH + 2GSH \rightarrow ROH + GSSG + H_2O$ reaction) rather than the direct reduction of RS. This is in agreement with previous yeast studies of Gasch et al. (2000) and Yoshimoto et al. (2002), who identified that Gpx-encoding mRNAs

do not exhibit significant changes in stressed yeast *S. cerevisiae* (majority of identified fold-changes were <= 1.5) (508,523).

In summary, the fold change increase LEV-2 SOD, IDH and ADH enzymes exposed to UV-B induced stress suggest that this yeast responds to UV-B by increasing the production of enzymatic antioxidants. This fold change increase co-occurred with the reduction in yeast cell death rate (Figure 5.3) and with the increase in antioxidant activity, as measured by DPPH free radical quenching (Figure 5.2), and followed the fold change increase in bZip protein LEV_2_XP_007274754, which shares sequence similarity to yeast *S. cerevisiae* bZip transcription factor Yap1 and the animal bZip transcription factor Nrf2. This suggests that the *Sporobolomyces* yeast LEV-2 response to UV-induced oxidative stress is evolutionary conserved between yeast *S. cerevisiae* and animals, but further, phylogenetic, studies are required to describe the evolutionary relationship between enzymatic antioxidants of yeast LEV-2 and enzymatic antioxidant systems of other yeasts (such as *S. cerevisiae*) and animals.

5.5.2.5 DNA repair and replication

UV radiation is a genotoxic environmental agent that inflicts DNA damage by causing the oxidative stress and by inflicting direct damage to the DNA macromolecule. UV-A and UV-B light induce the cellular production of reactive oxygen-derived species, such as H_2O_2 (39,40), and the evidence for UV-induced oxidative damage has been found in algae (559), plants (560), animals (39,557) and bacteria (504). UV-induced oxidative stress is also considered to be a major source of UV-induced damage in yeasts (569). During oxidative stress, the highly reactive hydroxyl radicals (OH•) react with the DNA molecule and cause the formation of single and double stranded DNA breaks, base modifications and cross-linkage with proteins (5). In addition to inducing oxidative stress, UV-B light is also absorbed by DNA pyrimidine bases, thymine and cytosine, and induces photoreactions that lead to formation of mutagenic DNA photoproducts, such as cyclobutane–pyrimidine dimers (CPDs) and 6–4 photoproducts (570), as well as formation of single stranded DNA breaks (37).

The maintenance of DNA is essential for all forms of life, and DNA repair mechanisms are well conserved across all domains of life, including bacteria, yeasts and animals (571). The photoreactivation mechanism that utilizes photolyase enzymes to repair CPDs and 6–4 photoproducts is considered to be the oldest and the simplest mechanism of DNA repair, as evinced by existence of photolyases in all domains of life, including archaea, and by the fact that it only utilizes a single enzyme (572). Contrasted to photoreactivation are excision repair mechanisms that do not repair the DNA damage, but instead remove the damaged section of DNA and replace it with

newly synthesized nucleotides (569). These mechanisms comprise three major categories: base excision repair (BER), which repairs small changes in DNA that do not alter the DNA helix structure; nucleotide excision repair (NER) which replaces "bulky" DNA adducts such as thymine dimers and 6,4-photoproducts; and mismatch repair mechanisms that repair erroneous insertions, deletions and incorporations of nucleic bases during DNA replication and recombination (571,573). Unlike the photoreactivation which utilizes only a single enzyme, excision repair pathways are comprised of numerous enzymes. For example, the BER pathway utilizes DNA glycosylase to recognize DNA damage and remove damaged nucleic base; apurinic/apyrimidinic (AP) endonuclease, AP lyase and phosphodiesterase to excise the deoxyribose phosphate residue left-over after removal of damaged nucleic base; DNA polymerase to repair the introduced DNA gap; and DNA ligase to connect the newly synthesized DNA with the rest of the DNA strand. The NER mechanism is also comprised of multiple enzymes including proteins that recognize DNA damage, endonucleases, helicases, ligases and other proteins required for regulation of the process (569). Notably, not all organisms possess all of the described DNA repair mechanisms; for example, photolyases required for photoreactivation repair of CPDs have been reported in bacteria, fungi, plants, invertebrates and many vertebrates, but not in humans, while photolyases that reverse 6–4 photoproducts have been found in fly *Drosophila*, frog *Xenopus laevis*, and certain snakes, but not in *E. coli*, yeast *S. cerevisiae* or humans (569).

Six yeast enzymes involved in the repair and replication of DNA were identified and quantified in this study (Figure 5.9). Identified proteins included DNA polymerases, DNA helicase, exonuclease and DNA ligase, all of which are involved in repair of single strand DNA breaks by BER and in the repair of double-strand breaks by non-homologous end-joining (574,575). In addition, one dUTP pyrophosphatase enzyme was identified and quantified; this enzyme is essential for DNA replication and repair as it prevents erroneous incorporation of uracil into the DNA (576). The expression levels of these enzymes were increased in the yeast LEV-2 sample exposed to 24 hours of UV irradiation (Figure 5.9), indicating that long-term UV-induced stress activates DNA repair in yeast LEV-2. This increase in expression of DNA repair enzymes correlated with the reduction in yeast cell death (Figure 5.3), suggesting that the yeast LEV-2 adapts to UV-induced stress by increasing the rate of DNA repair.

Notably, photolyase enzymes were not identified in this study, possibly because these proteins have high molecular mass and charge (for example, the yeast *S. cerevisiae* photolyase encoded by *PHR1* gene is a 66 kDa protein with pI value of 9.2), and it was demonstrated that MudPIT detection rates are low for large, charged proteins (577).

5.5.2.6 Heat-shock proteins and related proteins

Heat shock proteins (HSPs) and chaperones are highly conserved across all domains of life (578). HSPs are mainly involved in resistance to heat-induced stress, while chaperones also play a role in protein folding and unfolding, assembly of protein complexes, protein transport, cell-cycle control and protection against apoptosis (579). The MudPIT analysis of LEV-2 yeast exposed to long-term UV irradiation identified and quantified 24 proteins annotated as heat-shock proteins or chaperones (Figure 5.10). The expression of the majority of HSPs was increased in samples exposed to 24 hours of UV-B, co-occurring with the increase in expression of other stress resistance proteins (Figure 5.4) and reduction in cell death (Figure 5.1). This increase in expression was particularly noticeable for 10 kDa mitochondrial HSPs, Chaperonin ATPases TCP-1 and 90 kDa HSP proteins, indicating these families of HSP proteins play a major part of stress response in yeast LEV-2. This is possibly because the 90 kDa HSPs are involved in cellular signalling in addition to the function in stress response (580), and the TCP-1 chaperonins also mediate the ATP-dependent renaturation of proteins to assist in repair of UV and oxidative damage (581).

Interestingly, the expression levels of multiple HSP proteins, classified as 60 kDa HSPs and 70 kDa HSPs, were reduced in the LEV-2 samples exposed to 1 hour - 4 hours of UV-B (Figure 5.10). These results match a previous study of Gasch et al. (2000) that measured mRNA levels in yeast *S. cerevisiae* exposed to different sources of stress, and identified a significant increase in levels of mRNAs encoding 12 kDa, 30 kDa, 42 kDa and 104 kDa HSPs during heat shock, while levels of mRNAs encoding other HSPs were largely unchanged (508). The related study of Yoshimoto et al. (2002), measured mRNA levels of yeast *S. cerevisiae* under different environmental stresses (not including UV), and also identified that changes in levels of HSP-encoding mRNAs vary with the source and the duration of stress (508,523). The reduction in HSP levels observed for UV-exposed yeast LEV-2 is possibly due to the reduction in protein biosynthesis observed in samples exposed to 1 hour, 2 hours and 4 hours of UV-B (Figure 5.4) and indicates that 60 kDa HSPs and 70 kDa HSPs are unlikely to be a major part of UV response in this yeast.

5.5.2.7 Enzymes involved in MAA biosynthesis

Extremophile fungi inhabiting environments under high solar irradiance are known to produce UV-protective mycosporines such as mycosporine-glutamicol-glucoside (582,583), and certain freshwater *Rhodotorula* and *Cryptococcus* yeasts have been shown to produce UV-protective red pigments and fungal mycosporines when exposed to UV (584). The proteome of UV-tolerant yeast LEV-2 was analysed for enzymes involved in biosynthesis of mycosporines and

mycosporine-like amino acids (MAAs) to identify whether this yeast produces these compounds in response to UV-induced stress. Several enzymes involved in the shikimic acid pathway and biosynthesis of MAAs (Figure 5.12) were identified and quantified in the proteome of LEV-2 (Figure 5.11). These included enzymes involved in the biosynthesis of chorismate (chorismate mutase and chorismate synthase) and MAA biosynthesis enzymes 3-dehydroquinate synthase (DHQS), Transaldolase (TA) and hybrid NRPS-like synthase. However, multiple enzymes critical for MAA biosynthesis were not detected in this yeast (OMT, EVS, DHQS and DAHPS), and the expression levels of identified proteins were reduced in irradiated LEV-2 cultures (Figure 5.11), suggesting that yeast LEV-2 does not produce MAAs or mycosporines when exposed to UV-B radiation.

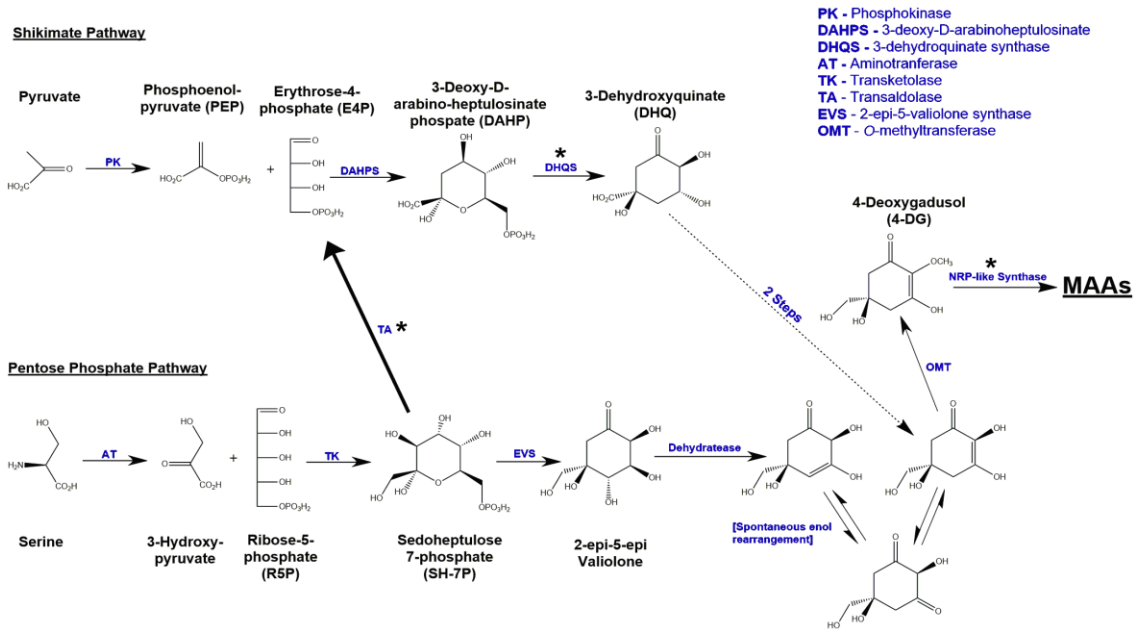


Figure 5.12 Postulated biosynthetic pathways to MAAs.

Mycosporine-like amino acid (MAA) biosynthesis involves shikimate pathway branching away from production of shikimic acid at DHQ, with DHQS playing critical role in synthesis of 4-deoxygadusol (4-DG). Alternate route of production is via pentose-phosphate pathway and involves multi-step conversion of SH-7P into 4-DG by EVS and OMT enzymes; Pathways converge at 4-DG, precursor of mycosporines and MAAs. Full arrows show experimentally validated reactions and enzymes, while dotted arrows show postulated reactions catalysed by one or more unknown enzymes. Enzymes detected in this study are marked with star. Figure adapted from (585).

To empirically verify the presence (or absence) of the MAAs, UV absorbance spectra were measured for the methanol extracts of cell pellets of UV-irradiated LEV-2 cultures and non-irradiated controls. The UV absorbance of irradiated LEV-2 cultures was considerably higher than the absorbance of non-irradiated controls (Figure 5.13), suggesting that yeast LEV-2 produces UV-absorbing compounds during UV-B exposure. The methanol extract of irradiated

LEV-2 cultures, however, was red-pigmented (as opposed to MAA extracts, which are colourless), the absorbance maximum of the extract was ~300 nm, which is different from known MAAs, and the preliminary HPLC analysis of the methanol extract of the irradiated LEV-2 culture did not detect the absorbance peak characteristic of known MAAs [data not shown]. These results suggested that the *Sporobolomyces sp.* yeast LEV-2 does not produce MAAs, but instead produces currently uncharacterized UV-absorbing red pigment. This is in accordance with previous research of Moliné et al. (2009) who found that yeasts producing carotenoid red pigments are highly tolerant to UV, when compared to unpigmented yeast strains (505).

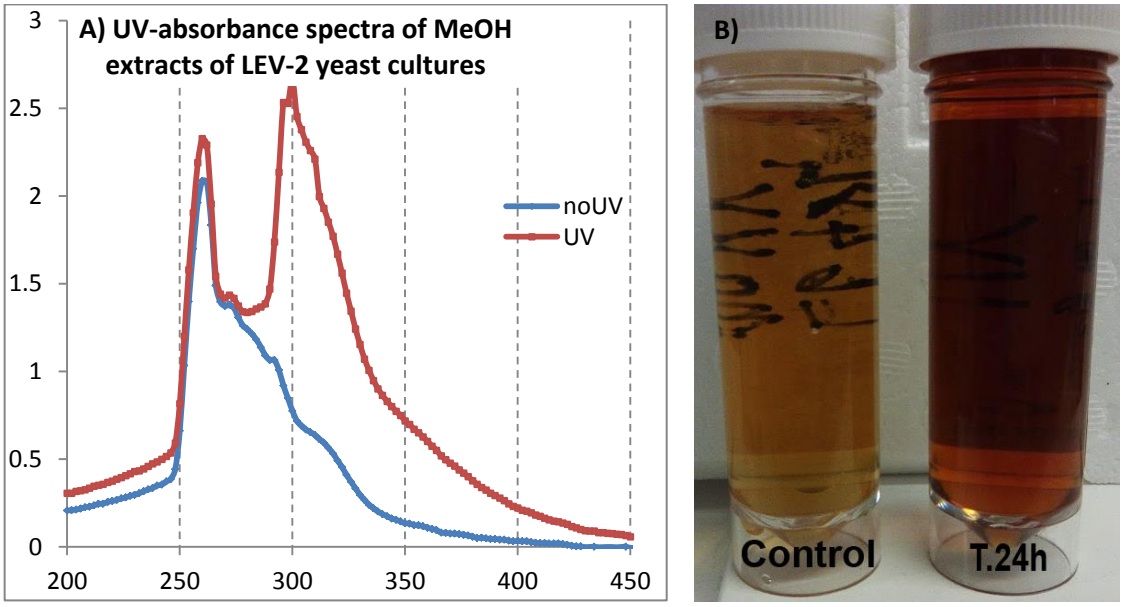


Figure 5.13 UV absorbance and pigment produced by yeast sample LEV-2

The figure displays UV-absorbance spectrum methanol extract of LEV-2 yeast culture cultivated without the UV exposure (blue) overlaid with the spectrum of extract of LEV-2 culture exposed to UV-B (red). The samples were cultivated for 24 hours and standardised to ~3 x 10^7 cells/mL. The curves present mean values of absorbance spectra of three experiments. The photograph of methanol extracts of UV-exposed culture (T.24h) and non-irradiated control (Control) is presented in figure B); the photograph has been cropped and no post-processing was performed.

5.5.3 Proposed model of yeast LEV-2 stress response

The results of the quantified proteome of LEV-2 exposed to extended UV-B irradiation, presented in section 5.4, were related to the antioxidant activity of cell-lysis extracts of LEV-2 cultures, measured by the DPPH assay (Figure 5.2), and to survival rates of LEV-2 cultures exposed to long-term UV-B irradiation (Figure 5.1) to construct the model of UV-stress response of the *Sporobolomyces* yeast LEV-2.

This comparison of yeast viability, antioxidant activity and protein fold changes of UV-exposed LEV-2 suggested that a moderately long UV-B irradiation (~1 hour) caused the fold change reduction in the yeast LEV-2 proteins involved in the carbohydrate metabolism, energy metabolism and protein biosynthesis (Figure 5.4), caused the moderate (~25%) reduction in the yeast viability (Figure 5.1), and triggered fold change increase in cellular signalling proteins related to Ras and Yap1 pathways (Figure 5.5, Figure 5.6). The continued irradiation further reduced the energy metabolism and viability of yeast cultures, with the maximal impact observed after 4 hours of UV-B. The expression level fold changes of proteins involved in yeast energy metabolism, however, increased after 8 hours of UV-B exposure and reverted to levels close to the control after 24 hours of UV-B exposure (Figure 5.4). This co-occurred with the reduction of yeast cell death, observed after 8 hours or longer UV-B exposure (Figure 5.3), and with an increase in expression levels of LEV-2 stress-response proteins such as enzymatic antioxidants (Figure 5.8) and certain heat-shock proteins (Figure 5.10). The levels of stress-response proteins also correlated with the free-radical quenching activity of the yeast cell extracts, as measured by the DPPH free radical scavenging assay (Figure 5.2).

These results can be best explained by proposing a yeast stress response and adaptation model, illustrated in the Figure 5.14. The model postulates that yeast UV response can be modelled in a four-step process: 1) UV exposure inflicts direct damage to the cell and induces oxidative stress, which in the short term induces bZip protein mediated signalling, analogous to Nrf2 signalling in animals or Yap1 signalling in the yeast *S. cerevisiae* (“signalling” phase); 2) The damage caused by the extended exposure to UV impairs carbohydrate and energy metabolism of the cell, and RS deplete cellular antioxidants (“stress” phase); 3) As the cellular defences are expressed, the yeast undergoes an “adaptation” phase where newly synthetized antioxidant enzymes (SODs, Cat, GTRx and NQR) and small molecule antioxidants (e.g. glutathione, ascorbic acid and α-tocopherol), DNA-repair enzymes, and UV-protective metabolites such as UV-absorbing carotenoids reduce the stress level and cellular metabolism starts recovery; 4) this leads to a “stress resistant” phase of growth where cellular metabolism and growth stabilize at a rate lower then pre-stress conditions.

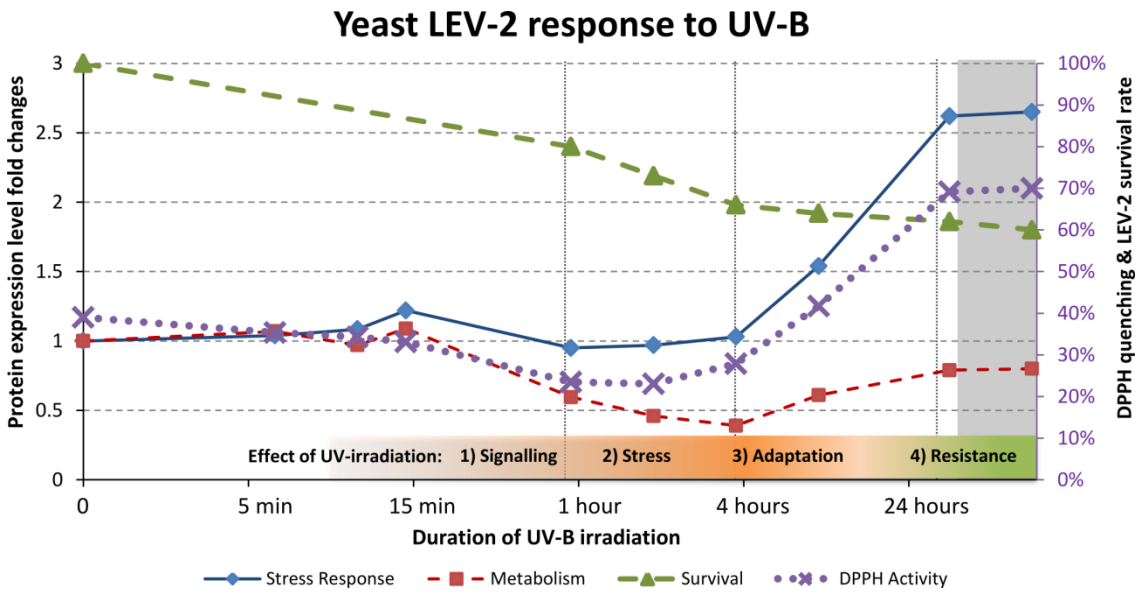


Figure 5.14 Proposed proteomics-based cellular response model for the *Sporobolomyces* yeast LEV-2 response to extended UV-B exposure

Figure presents the proposed model of stress response of UV-tolerant yeast LEV-2. The fold changes of proteins involved in stress response (blue line), and proteins involved in carbohydrate metabolism (red line) correspond to main (left side) Y-axis. The yeast survival curve (green line) and the DPPH quenching activity of yeast cell extracts (purple line) are expressed in percentages and correspond to the secondary (right side) Y-axis. The phases of stress response are denoted on the bottom on the chart. The chart is based on MudPIT experiment of LEV-2 yeast, and the grey shaded part of the chart represents the predicted patterns of yeast proteome, based on the "adaptation model" explained in the text.

The induction of stress resistance by low-intensity, sub-lethal, stress has been previously described in different eukaryotic models, including the yeast *S. cerevisiae* (586), mammalian cell cultures (587) and *in-vivo* animal models (469). For example, in a study by Davies et al. (1995), *S. cerevisiae* cultures conditioned by exposure to low concentration (0.4 mM) of H_2O_2 were found to survive, with ~90% viability rate, the subsequent exposure to high concentration (3 mM) of H_2O_2, lethal to unconditioned yeasts. The microarray studies of *S. cerevisiae* exposed to different environmental stresses, such as heat-shock, H_2O_2 and toxic metals, found also that induction of proteins involved in stress response is transient, and the expression levels revert to levels close to non-stressed yeast during the prolonged stress as yeast adapts to stress (507). Similar adaptation was observed for the fly *D. melanogaster* and for mouse cell cultures, where the exposure to a low dose of H_2O_2 conditioned the observed animals or cells to the following oxidative shock caused by high concentration of H_2O_2 (588). In animals, the adaptation to oxidative stress is mediated by the Nrf2 pathway; The low dose of oxidant, such as H_2O_2, induces the transcription of antioxidant and cytoprotective genes regulated by the bZip transcription factor Nrf2, and stimulates an increase in tolerance to oxidative stress (464).

Numerous studies of mouse models and animal cell lines have demonstrated that pre-treatment by Nrf2-activators such as SFN increased tolerance to oxidative stress in animals (156). For example, Nrf2 upregulation was found to protect human cells against cigarette smoke (93), UV-induced oxidative damage (225) and toxicity of drugs such as cisplatin (186). The activation of Nrf2 was also shown to protect mouse models against UV-induced carcinogenesis (469), while the Nrf2-knockout mouse models were shown to lack the ability to adapt to oxidative stress caused by carcinogens such as benzo[a]pyrene (184), and by drugs such as acetaminophen (185). bZip transcription factors have been shown to be evolutionary conserved between yeasts and animals (408,510), and *S. cerevisiae* bZip protein Yap1 regulates the response to oxidative stress in yeast (475), indicating that the molecular mechanism of adaptation to oxidative stress is evolutionary conserved between yeasts and animals.

In this study, the *Sporobolomyces* yeast LEV-2 bZip protein LEV-2_XP_007274754.1 showed a significant fold change increase in yeast cultures exposed to UV-B (Figure 5.5). The increase in expression of the bZip protein matched the results of previous studies that showed increase in expression of *S. cerevisiae* Yap1 protein in yeast exposed to oxidative stress (516) and the increase in bZip protein in animal cells exposed to oxidants (518,519). This suggests that the LEV-2 bZip protein LEV-2_XP_007274754.1 is a homolog of the *S. cerevisiae* transcription factor Yap1 and the vertebrate transcription factor Nrf2, and that the bZip protein mediated response to oxidative stress is conserved between LEV-2, *S. cerevisiae* and animals, as was previously predicted by our *in-silico* studies (408,510). It should be noted, however, that further studies are required to quantify the relative contributions of different signalling pathways in stress response of yeast LEV-2.

5.6 Conclusions and further research

The primary objective of this study was to determine if homologs of the vertebrate bZip protein Nrf2 play a role in stress response of UV-tolerant yeasts, such as carotenoid-producing yeasts of Genus *Sporobolomyces.* The quantitative MudPIT analysis of the *Sporobolomyces* yeast LEV-2 exposed to long term UV-B irradiation was followed by the functional annotation of the yeast LEV-2 proteome. This study identified four basic leucine zipper (bZip) proteins, and suggested that a bZip protein designated LEV-2_XP_007274754.1 is a homolog of the yeast bZip protein Yap1 and animal bZip protein Nrf2, and initiates the response to UV-induced oxidative stress in yeast LEV-2.

The secondary goal of this study was to describe the UV-stress response of yeast LEV-2, and evaluate if the stress response is conserved between LEV-2, other yeasts such as *S. cerevisiae* and animal models. A quantitative MudPIT proteomics analysis of LEV-2 cultures exposed to extended UV-B irradiation led to the proposal of a 4-step model, where 1) short-term UV irradiation induces cellular signalling; 2) depletion of cellular antioxidants and reduction in cellular metabolism; 3) adaptation by high expression of antioxidant enzymes and production of UV-absorbing red pigments; and 4) shift into stress-resistant phase of growth characterized by high expression of antioxidants and reduced metabolism. This model (Figure 5.14) matches the adaptation to stress observed also in *S. cerevisiae* and in animals, indicating that the molecular mechanisms of adaptation to oxidative stress are evolutionary conserved between *Basidiomycota* fungi, *Ascomycota* fungi, and animals.

Further, currently ongoing, research will focus on the sequencing and assembly of the genome of the yeast LEV-2 to facilitate phylogenetic studies of bZip proteins and antioxidant enzymes of LEV-2 and to enable the database matching of tandem mass spectra generated from LEV-2 proteins to kelch-like proteins encoded by this yeast. The DNA sequences of bZip proteins and kelch-like proteins of LEV-2 will allow for the design of primers for qPCR analysis to compare the mRNA levels of these proteins between the UV-exposed LEV-2 and a non-irradiated control. The yeast genome sequence will also be used to determine the protein sequences of bZip and kelch-domain proteins in LEV-2. This will enable the investigation of protein-protein interactions of these proteins in UV-stressed LEV-2 to investigate whether the yeast bZip proteins interact with kelch-like proteins, as observed in the animal Keap1-Nrf2 pathway.

Chapter 6: General discussion

6.1 Evolution of Keap1-Nrf2 pathway

The first hypothesis of this project was that the vertebrate basic leucine zipper (bZip) transcription factor Nrf2 has homologs in microorganisms, and that these homologs are involved in the regulation of microbial response to oxidative stress.

This hypothesis was tested by the phylogenetic analysis of the evolution of Nrf2 and its inhibitor Keap1. During the first part of this phylogenetic study, presented in chapter 2, a novel HMM-based pipeline for data mining of distant homologs was developed and utilized to identify the microbial homologs of Nrf2 and Keap1 in the GenBank and UniProt protein databases. The sequences encoding the homologs of vertebrate Nrf2 protein were identified in sequenced genomes of fungi and animals, but not in bacterial genomes, while the sequences encoding proteins related to human Keap1 protein were identified in genomes of organisms from all domains of life. Phylogenetic reconstructions of these sequences identified that the evolution of genes encoding homologs of Keap1 and Nrf2 proteins followed the putative evolution of major eukaryotic phyla, with the exception of sequences in nematode worms, which were grouped with sponges (for Keap1 protein sequences) or fungi (for Nrf2 protein sequences). These results indicated that the Nrf2 based signalling evolved prior to the divergence of fungi and Metazoa, and that homologs of vertebrate Nrf2-encoding genes exist in fungi and animals. The unexpected position of Genus *Caenorhabditis* implied that the Nrf2 signalling of nematode worms differs from the rest of animals (408), which is consistent with empirical studies that found that the SKN-1 protein, a *C. elegans* homolog of vertebrate Nrf2, does not interact with Kelch-like proteins (324), but is instead controlled by WDR-23 protein similar to human beta-transducin repeat-containing protein (β-TrCP) (436). The *C. elegans* WDR-23 and human β-TrCP proteins share structural topology and sequence similarity (408), and the vertebrate Nrf2 is known to be ubiquitinated and degraded in Keap1-independent fashion by β-TrCP:SCF E3 ubiquitin ligase complex (140), implying that nematodes possess Keap1-independent, but not Keap1-dependant mechanism of Nrf2 regulation. These results also suggest that *C. elegans* might not be a suitable model for studies of molecular mechanisms controlling the activity of Nrf2 in higher animals, and that the other invertebrates such as the fly *Drosophila melanogaster*, which shares Keap1-dependant control of Nrf2 (435), are more suitable invertebrate models for the studies of Keap1-Nrf2 interaction.

Based on the results presented in chapter 2, it was postulated that the evolution of Keap1-Nrf2 pathway might have been driven by the evolutionary pressures incurred by the rise of oxygen levels during the great oxygenation event (GOE). This assumption was tested by reconstructing the temporal framework of evolution of Nrf2 proteins, and by comparing the fossil-calibrated phylogenetic tree to geophysics models of the atmospheric oxygen levels over geological time periods. The results of this study (510), presented in the chapter 3, indicate that the time frame of major divergence points in evolution of Nrf2 co-occurs with the rise in oxygen levels in stages 4 and 5 of oxygenation of the atmosphere (405), during which the atmospheric concentration of oxygen started to rise towards the present levels. The analysis of evolutionary pressures on the Nrf2-encoding DNA sequences from microorganisms and animals indicated that the cyanobacterial Nrf2-like sequence (used as an outgroup) and fungal sequences were not under significant evolutionary pressure, presumably due to low oxygen levels in Precambrian era, while animal sequences were under strong selective pressure to retain the functional Nrf2 sequence, possibly caused by the pressure to retain the Nrf2-mediated stress response in the oxidizing environment of the Phanerozoic era. The examination of protein sequence of Nrf2 homolog of the worm *Caenorhabditis elegans* and the DNA sequence encoding it matched the results of our previous study (408), and indicated that *SKN-1* gene (*C. elegans* homolog of gene encoding the Nrf2 protein in humans), was not under significant selective pressure and that the time-frame for its divergence from the basal metazoan phylum did not match fossil records or the commonly accepted model of animal evolution (589). This was in contrast to all the other examined animal Nrf2-encoding genes which were found to be under selective pressure and matched the putative model of animal evolution

In summary, the research presented in chapters 2 and 3 of this thesis provided phylogenetic, *in-silico*, evidence to support the hypothesized existence of homologs of Nrf2 in fungi and all major animal phyla, and demonstrated the co-occurrence of major changes in atmospheric oxygen levels with the evolution of Nrf2 signalling. Phylogenetic reconstructions suggested that nematodes have markedly different control of Nrf2 signalling than other animals, and might not be a suitable model for the study of Keap1-dependant mechanisms of Nrf2 degradation in animals. These studies also suggested that basal metazoans, such as cnidarians or sponges, might possess the Keap1-dependant control of Nrf2 activity; this is in accordance with the previous bioinformatics study of the sea anemone *Nematostella vectensis* by Goldstone (2008), who identified that the genome of *N. vectensis* encodes the homologs of proteins involved in the Keap1-Nrf2 pathway (402), but these *in-silico* results are yet to be empirically confirmed.

6.2 Mycosporine-like amino acids for activation of Nrf2

The second hypothesis of this project was that some microorganisms produce small molecule secondary metabolites for endogenous control of a bZip protein mediated response to oxidative stress, and that such compounds have the potential to control the activity of Nrf2 in vertebrates. The mycosporine-like amino acids (MAAs) are a class of low molecular mass, UV-absorbing compounds, produced by a wide range of microorganisms (64). While primarily associated with UV protection (230), certain MAAs are also natural antioxidants (232), and have also been associated with protection against osmotic, desiccation and thermal stresses (64,231–234).

The research described in chapter 2 utilized the *in-silico* model of Keap1-Nrf2 protein-protein interaction (PPI) to evaluate if the MAAs have the potential to inhibit the ubiquitination and degradation of Nrf2 by binding to the Keap1 kelch domain β-propeller to disrupt the interaction between Keap1 and Nrf2 proteins. This structure based virtual screen identified that certain MAAs, such as mycosporine-glycine and porphyra-334, have the potential to bind to the human Keap1-Nrf2 binding pocket, and implied that MAAs could disrupt the CRL^{Keap1} protein complex to inhibit the degradation of Nrf2 and enhance the expression of proteins encoded by genes regulated by the transcription factor Nrf2.

The *in-silico* study was translated into an *in-vitro* model in the study presented in chapter 4. This study used the fluorescence polarization (FP) assay (159) and the thermal shift assay (294) to quantify the competitive inhibition of human Keap1-Nrf2 protein-protein interaction (PPI) by MAAs porphyra-334, shinorine and palythine. The Keap1-Nrf2 PPI was modelled *in vitro* using the human Keap1 Kelch-repeats β-propeller domain and the synthetic peptide FITC-β-DEETGEF-OH containing the ETGE motif with high affinity for the Keap1 β-propeller (135). The results of these assays demonstrated that MAAs porphyra-334 and shinorine are competitive inhibitors of Keap1-Nrf2 binding *in vitro*. While the activities of these compounds were not very high (with 50% inhibition of Keap1-peptide binding attained at concentration of ~100 μM), it is important to note that inhibition of the CRL^{Keap1} protein complex activity does not necessary require the inhibition of high affinity Keap1-(Nrf2-ETGE motif) binding, but can also be achieved by the inhibition of low affinity Keap1-(Nrf2-DLG motif) interaction (135). The low affinity, Keap1-(Nrf2-DLG motif), interaction was not investigated in the study presented in chapter 4, but as the MAAs interact with Keap1 β-propeller, it would be expected that MAAs also inhibit the Keap1-(Nrf2-DLG motif) interaction. That is because the DLG motif of Nrf2 interacts with the same binding pocket on the Keap1 β-propeller as the ETGE motif (135). The analysis of Nrf2-DLG and Nrf2-ETGE motif binding affinities for the Keap1

β-propeller was performed by Tong et al. (2006), and the study concluded that the affinity of Nrf2-ETGE motif for the Keap1 β-propeller is approximately 2 orders of magnitude higher than the affinity of Nrf2-DLG motif (590). Because of this, the MAAs porphyra-334 and shinorine have the potential to disrupt the activity of CRL^{Keap1} complex at concentration considerably lower than 100 μM. While the study of Nrf2 activation by MAAs *in vivo* was beyond the scope of this thesis, a recent study by Ryu et al. (2015) (463) reported that MAA porphyra-334 activated the Nrf2 signalling pathway in UV-A irradiated human cell model, which is in accordance with *in-silico* models and *in-vitro* studies presented in chapters 2 and 4. The study by Ryu et al. found that porphyra-334 is active in concentrations as low as 10μM, which is below the 100 μM used in our *in-vitro* experiments, and is consistent with the assumption that activation of Nrf2 by inhibition of Keap1 binding to Nrf2-DLG is achievable by porphyra-334 in concentration considerably lower than 100 μM.

In summary, the *in-silico* study presented in chapter 2 and the *in-vitro* experiments described in chapter 4 demonstrate that certain MAAs, produced by marine microorganisms and algae, have the potential activate the Nrf2 signalling in animals. This activity of MAAs is yet to be validated *in vivo*.

6.3 Investigation of a yeast model of Nrf2-mediated stress response

The third hypothesis of this research project was that a microorganism based Nrf2 activation assay can provide a feasible alternative to animal cell-based assays for pre-animal studies. This hypothesis was tested by examining, at the proteome level, the UV-B response of the UV-tolerant yeast *Sporobolomyces sp.* designated LEV-2. The study presented in the chapter 5 of this thesis used the quantitative MudPIT approach to quantify the expression of yeast proteins in LEV-2 yeast cultures exposed to UV-B radiation, ranging from no UV to 24 hours of exposure. The study identified 751 yeast proteins for which expression levels could be quantified for all samples. The fold changes of ~30% of quantified yeast proteins were significantly reduced in UV-irradiated samples exposed to 1 and 2 hours of UV-B, while the fold changes of ~37% of proteins were significantly increased in LEV-2 exposed to 24 hours of UV-B. The functional annotation of these proteins indicated that yeast metabolism and stress-response are inhibited after a moderately long (1 hour) exposure to UV, and this effect was confirmed by DPPH free radical quenching assay of yeast cell lysates. The long-term exposure to UV-B led to recovery of yeast metabolism and an increase in expression of LEV-2 stress-response proteins such as catalase, superoxide dismutases and glutathione peroxidases. This increase in stress response followed the increase in expression levels of a basic leucine zipper

(bZip) transcription factor of yeast LEV-2. These results indicated that the Nrf2-like bZip transcription factor is involved in oxidative stress response in LEV-2, and are in accordance with the results of phylogenetic studies presented in chapters 2 and 3, which identified Nrf2 homologs in fungi (408,510). In addition, the bZip protein Yap1 is known to play a role in the response to oxidative stress in the yeast *Saccharomyces cerevisiae* (330), and Yap1 shares a sequence similarity to the human Nrf2 protein (Appendix A-7). While this implies that the yeast LEV-2 possesses the stress response signalling similar to animal Nrf2-mediated stress response, no kelch-domain proteins were identified in the MudPIT analysis of LEV-2, possibly due to problems with protein extraction due to the large size of kelch-domain proteins and their association with the cytoskeleton (577). Unlike vertebrate Nrf2 and its homolog in the fly *Drosophila melanogaster*, the *S. cerevisiae* bZip transcription factor Yap1 has not been reported to be ubiquitinated and degraded by the kelch-protein associated complex (496), and the bioinformatics studies presented in chapters 2 and 3 suggested that fungi, in general, do not possess a Keap1-like inhibition of Nrf2. The protein-protein interactions of bZip proteins of LEV-2 are, however, currently unknown, and further research is required to determine if this yeast would be a suitable model for Keap1-Nrf2 interaction in vertebrates. In addition, the study presented in Chapter 5 utilised UV as source of oxidative stress and the UV radiation is also known to activate DNA-damage specific stress response pathways in animals as well as fungi (471). Furthermore, the discovered LEV-2 bZip transcription factor could also be functionally homologous to AP-1 bZip transcription factors involved specifically in UV response, and the sequence similarities could not be evaluated due to the lack of of genome sequence of yeast LEV-2. Thus the further research, possibly by exposing LEV-2 yeast to oxidants such as H_2O_2, is required to determine if bZip transcription factor discovered in yeast LEV-2 is involved in response to UV-induced DNA damage or specific response to oxidative stress.

In summary, the MudPIT analysis of UV-stress response of the *Sporobolomyces* yeast LEV-2 quantified the yeast response to UV-induced oxidative stress, led to the proposal of a stress response model for this UV-tolerant yeast and identified that a bZip transcription factor is likely involved in stress response of LEV-2. However, the validation (or refutal) of the hypothesis that stress-resistant yeast can provide an *in vivo* model of vertebrate Nrf2-based stress response requires further research to identify Keap1-like proteins in this yeast and to describe the protein-protein interactions of kelch-domain proteins and bZip transcription factors in LEV-2.

6.4 Conclusions

This thesis examined the evolution of the Keap1-Nrf2 pathway, the role of bZip proteins similar to the vertebrate transcription factor Nrf2 in yeast response to UV-induced oxidative stress, and assessed the potential for the activation of Nrf2 regulated genes by mycosporine-like amino acids. Based on the studies described in chapters 2 – 5 and discussed in this chapter, the conclusions of this thesis are the following:

1. The genomes of animals and certain fungi, such as UV-tolerant yeasts of the Genus *Sporobolomyces*, encode bZip transcription factors homologous to the human Nrf2 protein. Nrf2-mediated response to oxidative stress possibly evolved under the selective pressures incurred by the rise in levels of atmospheric oxygen over geological time.
2. The mycosporine-like amino acids, small compound natural products associated with taxonomically diverse marine organisms, are competitive inhibitors of human Keap1-Nrf2 interaction *in vitro.*

These conclusions provide a strategic platform for future research into the therapeutic benefits of microbial natural products that have evolved conjointly with the Keap1-Nrf2 antioxidant defence. Possible areas for future research are now described.

6.5 Future research

6.5.1 Phylogenetic study of animal antioxidant response elements

The genes under transcriptional control of the cis-promotor DNA motif known as antioxidant response element (ARE) have been studied in multiple human and mouse cell lines, and in mouse animal models (117), but ARE motifs have not been extensively studied in other animals. The assembled and annotated genomes of a large number of animals from all major animal phyla are currently available in the public domain, and a detailed bioinformatics study of ARE sequences in these organisms presents an opportunity to examine the genes under Nrf2 control in evolutionary distant animals, and to study the evolution of ARE motifs over geological time.

6.5.2 Network biology approach to evolution of Keap1-Nrf2 pathway

Multiple microarray, ChIP-Seq and proteomics studies were conducted on mouse models, and on mouse and human cell lines, to identify the genes transcriptionally activated by Nrf2 (125,128,129,180). Yet, the genes activated by Nrf2 have been less studied in other animal models, such as the fly *Drosophila melanogaster* and the worm *Caenorhabditis elegans*. Protein-protein interactions (PPIs) have been systematically characterized in these model organisms, and are deposited in PPI databases such as PRIDE (591) and STRING (592). The analysis of cellular PPI interaction networks of the Keap1 protein and the proteins encoded by Nrf2 activated genes would provide a comparison of biological functions regulated by Nrf2 in different model organisms, and allow the reconstruction of the evolution of Nrf2 regulated functional networks in animals. In addition, the analysis of these PPI networks would also allow the prediction of PPI hubs involved in the "dark side" effects of Nrf2 activation such as the skin pathologies observed in mouse models over-expressing Nrf2 protein (228,229).

6.5.3 Analysis of Keap1-Nrf2 interaction in the yeast LEV-2

The research presented in Chapter 5 of this thesis identified the yeast LEV-2 proteins involved in response to oxidative stress, compared the fold change patterns of these proteins relative to the duration of UV-B exposure, and identified that a bZip transcription factor is likely to activate the stress response in this yeast. However, the Keap1-like proteins of LEV-2 were not detected, and the protein-protein interactions (PPIs) of LEV-2 bZip transcription factors were not identified. Thus, further research is required to identify if Keap1-like proteins play a part in the control of stress response in LEV-2. Genome sequencing of this yeast would enable multiple experiments to examine kelch-domain proteins. For example, a sequenced genome would allow identification of potential Keap1-like proteins in this yeast and allow currently unassigned mass spectra to be matched to LEV-2 kelch domain proteins. In addition, *in-silico* translation of LEV-2 ORFs identified from the sequenced genome would allow a calculation of masses of potential bZip:kelch-domain protein complexes, which would facilitate a western blot based examination of such complexes in LEV-2. Finally, the knowledge of DNA sequences encoding the kelch-domain and bZip proteins in this yeast would enable the design of siRNAs for knock-down of expression of bZip proteins to examine the effect of bZip knockdown to LEV-2 stress response.

6.5.4 Examination of Keap1-Nrf2 interaction in basal metazoans

The Keap1-Nrf2 pathway is well described in vertebrates, the fly *Drosophila melanogaster* (435) and the nematode *Caenorhabditis elegans* (436). The genome of the basal metazoan *Nematostella vectensis* encodes homologs of vertebrate Nrf2 and Keap1 proteins (402,408), and the expression of the Nrf2 homolog of the cnidarian *Hydra magnipapillata* was found to be increased in injured hydra (593), but Keap1-Nrf2 protein-protein interaction was not studied in the basal metazoans. Several basal metazoan model organisms such as *Hydra magnipapillata, Nematostella vectensis* and *Hydractinia echinata* are suitable for cultivation in the laboratory (594,595) and genomes of these animals have been assembled (596–598), making them suitable models for study of Keap1-Nrf2 interactions. A study of PPI of Keap1-Nrf2 in a cnidarian model, for example by immunoprecipitation with an anti-Keap1 antibody, followed by western blotting with an anti-Nrf2 antibody, would elucidate if these proteins interact in cnidarian models. In addition, quantitative proteomics or microarray based quantification of gene expression in the cnidarian model subjected to Nrf2 and Keap1 knockdowns by siRNAs, and compared to unstressed and stressed animal controls, could be used to elucidate the genes under transcriptional regulation by Nrf2 in these animals. As well as describing the Nrf2 pathway in early animals, these studies would also identify whether nematodes such as *C. elegans* are an unusual case of Nrf2 evolution and have lost Keap1-dependant regulation, or if the Keap1-dependant regulation evolved after the nematode divergence from the basal metazoan phylum. If the former is the case, and the stress response in *C. elegans* is an exception to stress response observed in all other animals, it would imply that this worm might not be a suitable model for studies of ageing and stress in animals.

6.5.5 Further studies of MAA-induced activation of Nrf2-controlled genes

Over 20 MAAs have been identified and described, but only porphyra-334 has been studied for Nrf2 activation in cell culture models (463), while shinorine, porphyra-334 and palythine have been tested for inhibition of Keap1-Nrf2 interaction *in vitro* (chapter 4). MAAs have similar biophysical properties and chemical structures, and certain MAAs, such as mycosporine-glycine, scored highly in the *in-silico* model of human Keap1 antagonism (chapter 2) and are known to be potent *in-vitro* antioxidants (232,250). The mycosporine-glycine, however, was not tested in this project due to the difficulties in obtaining the purified compound and presents an opportunity for future *in vitro* and *in vivo* studies.

Appendix A

Appendix A-1: Sequences used in Keap1 and Nrf2 distant homology search

Appendix A-1 is included in electronic format on a supplied supplementary data files disk and is available online at http://www.sciencedirect.com/science/article/pii/S089158491500283X.

Appendix A-2: Distant homology to Keap1

DHSP (Appendix A-2) was used to search custom databases of archaeal, bacterial, fungal, plant and protozoan protein sequences for distant homology to Keap1. Results are shown in Table A-2.1.

Table A-2.1. Distant homologs of Keap1.

Table lists the predicted number of distant homologs of Keap1 protein detected by DHSP.

Database	Detected Keap1 homologs
Archaeal	45
Bacterial	~1200
Fungal	~2400
Plant	~5800

Appendix A-3: Distant homology to Nrf2

DHSP was used to search custom databases of archaeal, bacterial, fungal, plant and protozoan protein sequences for distant homology to Nrf2. Results are shown in Table A-3.1.

Table A-3.1. Distant homologs of Nrf2

Table lists the predicted number of distant homologs of Nrf2 protein detected by DHSP.

Database	Detected Nrf2 homologs
Archaeal	0
Bacterial	24
Fungal	68
Plant	14

The search detected no Nrf2 homology in the archaeal database and very limited homology in the plant and bacterial databases. No bacterial and plant taxa with homology to both Nrf2 and Keap1 proteins could be identified. Further analysis of the conserved domain sequences of Nrf2 (Neh1-Neh6) detected high homology to Neh1 in fungal taxa. Results are shown in Table A-3.2.

Table A-3.2. Taxonomic distribution of fungal Neh1 domain homologs of human Nrf2. Numbers of detected homologs are listed according to appropriate taxonomical levels.

KINGDOM	matches	PHYLUM	matches	CLASS	matches
Fungi	887	Undefined	36	Undefined	36
		Ascomycota	658	Dothideomycetes	92
				Eurotiomycetes	197
				Leotiomycetes	23
				Orbiliomycetes	7
				Pezizomycetes	7
				Saccharomycetes	117
				Schizosaccharomycetes	6
				Sordariomycetes	207
				Taphrinomycetes	2
		Basidiomycota	169	Agaricomycetes	98
				Dacrymycetes	5
				Exobasidiomycetes	3
				Tremellomycetes	21
				Ustilaginomycetes	17
				Wallemiomycetes	9
				Undefined	16
		Chytridiomycota	6	Chytridiomycetes	6
		Glomeromycota	8	Glomeromycetes	8
		Microsporidia	10	Undefined	10

Appendix A-4: Human Nrf2 and Keap1 homology in *Saccharomyces cerevisiae*

Saccharomyces cerevisiae human Nrf2 homologs:

Whole length Nrf2 homology: No homology detected

Nrf2 Neh1 domain homology

>gi|6322153|ref|NP_012228.1|__eV{1.2e-03}Ve__ Cst6p [Saccharomyces cerevisiae S288c] Basic leucine zipper (bZIP) transcription factor in ATF/CREB family; mediates transcriptional activation of NCE103 (encoding carbonic anhydrase) in response to low CO2 levels such as in the ambient air; proposed to be a regulator of oleate responsive genes; involved in utilization of non-optimal carbon sources and chromosome stability; relocalizes to the cytosol in response to hypoxia; CST6 has a paralog, ACA1, that arose from the whole genome duplication
MFTGQEYHSVDSNSNKQKDNNKRGIDDTSKILNNKIPHSVSDTSAAATTTSTMNNSALSRSLDPTDINYSTNMAGVVD
QIHDYTTSNRNSLTPQYSIAAGNVNSHDRVVKPSANSNYQQAAYLRQQQQQDQRQQSPSMKTEEESQLYGDILMNSGV
VQDMHQNLATHTNLSQLSSTRKSAPNDSTTAPTNASNIANTASVNKQMYFMNMNMNNNPHALNDPSILETLSPFFQPF
GVDVAHLPMTNPPIFQSSLPGCDEPIRRRRISISNGQISQLGEDIETLENLHNTQPPPMPNFHNYNGLSQTRNVSNKP
VFNQAVPVSSIPQYNAKKVINPTKDSALGDQSVIYSKSQQRNFVNAPSKNTPAESISDLEGMTTFAPTTGGENRGKSA
LRESHSNPSFTPKSQGSHLNLAANTQGNPIPGTTAWKRARLLERNRIAASKCRQRKKVAQLQLQKEFNEIKDENRILL
KKLNYYEKLISKFKKFSKIHLREHEKLNKDSDNNVNGTNSSNKNESMTVDSLKIIEELLMIDSDVTEVDKDTGKIIAI
KHEPYSQRFGSDTDDDDIDLKPVEGGKDPDNQSLPNSEKIK

>gi|398366073|ref|NP_011704.3|__eV{6.3e-03}Ve__ Bub1p [Saccharomyces cerevisiae S288c]
Protein kinase involved in the cell cycle checkpoint into anaphase; in complex with Mad1p and Bub3p, prevents progression into anaphase in presence of spindle damage; Cdc28p-mediated phosphorylation at Bub1p-T566 is important for degradation in anaphase and adaptation of checkpoint to prolonged mitotic arrest; associates with centromere DNA via Skp1p; involved in Sgo1p relocalization in response to sister kinetochore tension; paralog MAD3 arose from whole genome duplication"
MNLDLGSTVRGYESDKDTFPQSKGVSSSQKEQHSQLNQTKIAYEQRLLNDLEDMDDPLDLFLDYMIWISTSYIEVDSE
SGQEVLRSTMERCLIYIQDMETYRNDPRFLKIWIWYINLFLSNNFHESENTFKYMFNKGIGTKLSLFYEEFSKLLENA
QFFLEAKVLLELGAENNCRPYNRLLRSLSNYEDRLREMNIVENQNSVPDSRERLKGRLIYRTAPFFIRKFLTSSLMTD
DKENRANLNSNVGVGKSAPNVYQDSIVVADFKSETERLNLNSSKQPSNQRLKNGNKKTSIYADQKQSNNPVYKLINTP
GRKPERIVFNFNLIYPENDEEFNTEEILAMIKGLYKVQRRGKKHTEDYTSDKNRKKRKLDVLVERRQDLPSSQPPVVP
KSTRIEVFKDDDNPSQSTHHKNTQVQVQTTTSILPLKPVVDGNLAHETPVKPSLTSNASRSPTVTAFSKDAINEVFSM
FNQHYSTPGALLDGDDTTTSKFNVFENFTQEFTAKNIEDLTEVKDPKQETVSQQTTSTNETNDRYERLSNSSTRPEKA
DYMTPIKETTETDVVPIIQTPKEQIRTEDKKSGDNTETQTQLTSTTIQSSPFLTQPEPQAEKLLQTAEHSEKSKEHYP
TIIPPFTKIKNQPPVIIENPLSNNLRAKFLSEISPPLFQYNTFYNYNQELKMSSLLKKIHRVSRNENKNPIVDFKKTG
DLYCIRGELGEGGYATVYLAESSQGHLRALKVEKPASVWEYYIMSQVEFRLRKSTILKSIINASALHLFLDESYLVLN
YASQGTVLDLINLQREKAIDGNGIMDEYLCMFITVELMKVLEKIHEVGIIHGDLKPDNCMIRLEKPGEPLGAHYMRNG
EDGWENKGIYLIDFGRSFDMTLLPPGTKFKSNWKADQQDCWEMRAGKPWSYEADYYGLAGVIHSMLFGKFIETIQLQN
GRCKLKNPFKRYWKKEIWGVIFDLLLNSGQASNQALPMTEKIVEIRNLIESHLEQHAENHLRNVILSIEEELSHFQYK
GKPSRRF

>gi|6323696|ref|NP_013767.1|__eV{3.0e-03}Ve__ Far3p [Saccharomyces cerevisiae S288c] hypothetical protein; involved in recovery from cell cycle arrest in response to pheromone, in a Far1p-independent pathway; interacts with Far7p, Far8p,Far9p, Far10p, and Far11p; localizes to the endoplasmic reticulum; protein abundance increases in response to DNA replication stress
MNSGGSDSFDYLLQLTKALSAECRANRQETDRIELLLKRLAKQSGISYDNLSKNIIPDSWKDNASQKASPPTEAQKLI
SENFKLIYEIEKQEYFNTKAVALINNINEHFSYIKNFIDEQNAIRERNIATFSSEKLDERNKSLQQNYESLKTENEET
KKKLHSIIKQFEKLLKEVDWDRISKDSRDYSRFKKQLEYLQDTYQVLK

Human Keap1 homology:

Keap1 BTB domain homology:

>gi|6325211|ref|NP_015279.1|__eV{1.5e-02}Ve__ elongin C [Saccharomyces cerevisiae S288c]
MSQDFVTLVSKDDKEYEISRSAAMISPTLKAMIEGPFRESKGRIELKQFDSHILEKAVEYLNYNLKYSGVSEDDDEIP
EFEIPTEMSLELLLAADYLSI

Kelch beta propeller homology:

>gi|6321952|ref|NP_012028.1|__eV{1.6e-05}Ve__ Kel1p [Saccharomyces cerevisiae S288c] Protein required for proper cell fusion and cell morphology; functions in a complex with Kel2p to negatively regulate mitotic exit, interacts with Tem1p and Lte1p; localizes to regions of polarized growth; potential Cdc28p substrate; KEL1 has a paralog, KEL2, that arose from the whole genome duplication; is kelch repeat containing protein
MAGFSFAKKFTHKKHGKTPSDASISDQSREASLSTPPNEKFFTKQETPQKGRQFSQGYHSNVNKTSSPPMFARKQVSE
SRIQPSAVPPQQRNVSGPSTTLHKQLSKQREYTVWNRIKLQNSPFPRYRHVASAYVTDKNQIYVIGGLHDQSVYGDTW
ILTAFDNATRFSTTTIDISEATPPPRVGHAAVLCGNAFVVFGGDTHKVNKEGLMDDDIYLLNINSYKWTVPAPVGPRP
LGRYGHKISIIATTQMKTKLYVFGGQFDDTYFNDLAVYDLSSFRRPDSHWEFLKPRTFTPPPITNFTMISYDSKLWVF
GGDTLQGLVNDVFMYDPAINDWFIIDTTGEKPPPVQEHATVVYNDLMCVVGGKDEHDAYLNSVYFLNLKSRKWFKLPV
FTAGIPQGRSGHSLTLLKNDKILIMGGDKFDYARVEEYDLHTSDIDMQRGTIVYTLDLARIKDLCPGVMDVPTDTPTP
RNGNLDLATPVTPTSHQTKNMNVPISAAPLASAPSPAPKDFSDADRLNREVHNRNVSTEHQNQSHPVNSESHLIAEPN
ILTPYVPSESSQTPVMKITSNKPFDTPTIQKEPDLSETMDPTVGNQRIPSSIYGDNLTPANQIKNNSPILETLPSNEI
KTPQNGNIEEIKHLPDADEKIDSTTTFDQEINGDKLGTSSMSKVEEDGNVADEDDEIGVAQMASSPSKDQFKIKHYNE
SSELSQNNTEIDKLSEPVDITIKKSDTAGHDSANHVIDASDEKNVSPMGDVPTDTKNEEASVPINRDATTEVVDRALF
EKLRSELQSLKELTHEKALEAGAHIKELETELWQLKSQKNSGTTKEIDELDSVRLQSKCEILEADNHSLEDKVNELEE
LVNSKFLDIENLNEVIQFQNEKIKSLELEPNYKEKLEELQIEHENLSRENERLKNESKQHNEDIINNVANYSSQLGSL
ISHWKENRANSSFLESSSSLISVSDENGEKTVGEPYGDQSRHHRVVINKLTNRLDDLLERSQELTISKEKLSSEYHAL
KMEHSSLSQDVLVKENEIKKIQNDYKESISSMDSASKALMVSQRELEKYKSLNKKLIDELDELKFKNGVCSENFENGL
RSTEESSNNVKNSNSIRENQFNIKINDLKAELFITNQERDDLKSEVLELKKRLLNLENNTKQVNEDADSDLL

Whole length Keap1 homology:

>gi|398364545|ref|NP_012265.3|__eV{2.9e-04}Ve__ hypothetical protein YIL001W [Saccharomyces cerevisiae S288c] BTB containing, ankyrin repeats containing
MADKLMDKNFEELCYSCRTGDMDNVDRLISTGVNVNSVDKFDNSPLFLASLCGHEAVVKLLLQRGAVCDRDRYEGARC
IYGALTDTIRDTLLSYDISKAVDVKQPFATHISSMYNDEGFLKRDITFRVSNGKLFTAHKFLLCARSEILAEKMVNEW
AKHEIVSLEVRPDIFDIFLKFLYLIPILHQIEPGQYEELIELSSKFDIELLPEFLDKARHTADPTEKSRLMSDYQYKF
TEVARSQLLIFVNNCIFRSTVDLANSERRVFSLMNCPAYPDVQLMVKNRNGAIRIYPCHLAVLSRAEYFKVMFTNNFK
EKVTYIKAKHVTGKYNSIIPQLTLPNCEFEVAEIILRYLYADNTDIPWMYAVDVLLLADILLEDRLKTIASTIITQSK
EFIQQYNVFDVLYLSWEIGVERLEQFAAKFIAIHLQELYKDPEIKRAIMLSSQRISLRQETDTIELVDDIRYYLLRKY
SFEPDDVELFENQDDLEYLKQVGYLEYRKDMGMLDNILADLELDV

>gi|6321952|ref|NP_012028.1|__eV{2.3e-05}Ve__ Kel1p [Saccharomyces cerevisiae S288c]
MAGFSFAKKFTHKKHGKTPSDASISDQSREASLSTPPNEKFFTKQETPQKGRQFSQGYHSNVNKTSSPPMFARKQVSE
SRIQPSAVPPQQRNVSGPSTTLHKQLSKQREYTVWNRIKLQNSPFPRYRHVASAYVTDKNQIYVIGGLHDQSVYGDTW
ILTAFDNATRFSTTTIDISEATPPPRVGHAAVLCGNAFVVFGGDTHKVNKEGLMDDDIYLLNINSYKWTVPAPVGPRP
LGRYGHKISIIATTQMKTKLYVFGGQFDDTYFNDLAVYDLSSFRRPDSHWEFLKPRTFTPPPITNFTMISYDSKLWVF
GGDTLQGLVNDVFMYDPAINDWFIIDTTGEKPPPVQEHATVVYNDLMCVVGGKDEHDAYLNSVYFLNLKSRKWFKLPV
FTAGIPQGRSGHSLTLLKNDKILIMGGDKFDYARVEEYDLHTSDIDMQRGTIVYTLDLARIKDLCPGVMDVPTDTPTP
RNGNLDLATPVTPTSHQTKNMNVPISAAPLASAPSPAPKDFSDADRLNREVHNRNVSTEHQNQSHPVNSESHLIAEPN
ILTPYVPSESSQTPVMKITSNKPFDTPTIQKEPDLSETMDPTVGNQRIPSSIYGDNLTPANQIKNNSPILETLPSNEI
KTPQNGNIEEIKHLPDADEKIDSTTTFDQEINGDKLGTSSMSKVEEDGNVADEDDEIGVAQMASSPSKDQFKIKHYNE
SSELSQNNTEIDKLSEPVDITIKKSDTAGHDSANHVIDASDEKNVSPMGDVPTDTKNEEASVPINRDATTEVVDRALF

EKLRSELQSLKELTHEKALEAGAHIKELETELWQLKSQKNSGTTKEIDELDSVRLQSKCEILEADNHSLEDKVNELEE
LVNSKFLDIENLNEVIQFQNEKIKSLELEPNYKEKLEELQIEHENLSRENERLKNESKQHNEDIINNVANYSSQLGSL
ISHWKENRANSSFLESSSSLISVSDENGEKTVGEPYGDQSRHHRVVINKLTNRLDDLLERSQELTISKEKLSSEYHAL
KMEHSSLSQDVLVKENEIKKIQNDYKESISSMDSASKALMVSQRELEKYKSLNKKLIDELDELKFKNGVCSENFENGL
RSTEESSNNVKNSNSIRENQFNIKINDLKAELFITNQERDDLKSEVLELKKRLLNLENNTKQVNEDADSDLL

Alignment of human Keap1 and *S. cerevisiae* Skn7:

Max Score 17.7 Total Score 67.0 Cover 11% E-value 1.6 Identity 38%

Alignment statistics for match #1

Score 17.7; E-value 1.6; Identities 12/32(38%); Positives 14/32 (43%); Gaps 4/32 (12%)

```
Query 378  GGRNNSPDGNTDSSALDCY--NPMTNQWSPCA  407
           G RN  P GNT+ +       N  TN  SP
Sbjct 317  GNRN--PTGNTNPATTTAIQSNNNTNNASPAT  346
```

Alignment statistics for match #2

Score 16.9; E-value 2.9; Identites 8/15(53%); Positives 9/15(60%); Gaps 0/15(0%)

```
Query  337  QSLSYLEAYNPSDGT  351
            QSL+ L   NPS  T
Sbjct  529  QSLAMLPQDNPSTTT  543
```

Alignment statistics for match #3

Score 16.9; E-value 3.0; Identities 5/8(63%); Positives 6/8(75%); Gaps 0/8(0%)

```
Query  586  PDTDTWSE  593
            PD  TW+E
Sbjct  101  PDIVTWTE  108
```

Alignment of human Keap1 and *S. cerevisiae* YAP1

Alignment statistics for match #1

Score 18.9; E-value 0.65; Identities 11/30(37%); Positives 12/30(40%); Gaps 0/30(0%)

```
Query  504  NTIRSGAGVCVLHNCIYAAGGYDGQDQLNS  533
            NT  S   +  L N IY      G D  NS
Sbjct  246  NTPNSSTSMDWLDNVIYTNRFVSGDDGSNS  275
```

Alignment of human Nrf2 and *S. cerevisiae* YAP1

Max score 24.6 Total 109 Query Cover 27% E-value 0.011 Identity 27%

Alignment statistics for match #1

Score 24.6; E-value 0.011; Identities 17/63(27%); Positives 32/63(50%); Gaps 7/63(11%)

```
Query  486  RRRGKNKVAAQNCRKRKLENIVELEQDLDHLKDEKEKLLKEKGENDKSLHLLKKQLSTLY  545
            +R  +N+ A +  R+RK   + ELE+ +  L+  ++       +N+     L+ QL TL
Sbjct  69   KRTAQNRAAQRAFRERKERKMKELEKKVQSLESIQQ-------QNEVEATFLRDQLITLV  121

Query  546  LEV  548
             E+
Sbjct  122  NEL  124
```

Alignment statistics for match #2

Score 20.0; E-value 0.33; Identities 11/43(26%); Positives 21/43(48%); Gaps 2/43(4%)

```
Query  449  HIPFPVEKIINLPVVDFNEMMSKEQFNEAQLALIRDIRRRGKN  491
            ++ FP +   N+ +  F+E  S+ +F+       RD  + G N
Sbjct  469  NLAFPDDNSTNIQLQPFSESQSQNKFDYDM--FFRDSSKEGNN  509
```

Alignment statistics for match #3

Score 17.7; E-value 1.6; Identities 13/40(33%); Positives 18/40(45%); Gaps 2/40(5%)

```
Query  16   VSREVFDFSQRRKEYELEKQKKLEKERQEQLQKEQEKAFF  55
             ++  F   + RK  ELE  KK++     Q Q E E  F
Sbjct  76   AAQRAFRERKERKMKELE--KKVQSLESIQQQNEVEATFL  113
```

Alignment of Human Nrf2 and *S. cerevisiae* SKN7

Max score 16.9 Total score 49.3 Cover 11% E-value 2.6 Identity 38%

Alignment statistics for match #1

Score 16.9; E-value 2.6; Identities 5/13(38%); Positives 8/13(61%); Gaps 0/13(0%)

```
Query  407  ENTPEKELPVSPG  419
            ++ P    PV+PG
Sbjct  536  QDNPSTTTPVTPG  548
```

Alignment statistics for match #2

Score Expect Method Identities Positives Gaps

Score 16.5; E-value 3.5; Identities 10/37(27%); Positives 20/37(54%); Gaps 1/37(2%)

```
Query  16   VSREVFDFSQRRKEYELEKQKKLEKERQEQLQKEQEK  52
            VS++ F  + RR+  +L+K+  + K      + E +K
Sbjct  235  VSKDAFG-NLRRRVDKLQKELDMSKMESYATKVELQK  270
```

Alignment of *C. elegans* SKN1 and *S. cerevisiae* SKN7

No significant similarity found

Alignment of *S. cerevisiae* YAP1 and *C. elegans* SKN1

Score 17.3 Max score 83.5 Coverage 13% E-value 2.0 Identity 33%

Alignment statistics for match #1

Score 17.3; E-value 2.0; Identities 5/15(33%); Positives 9/15(60%); Gaps 0/15(0%)

```
Query  132  RNDSKVLEYLARRDP  146
            R+D +++EY     P
Sbjct  133  RDDQRMMEYFMSNGP  147
```

Alignment statistics for match #2

```
Score 17.3; E-value 2.2; Identities 6/11(55%); Positives 8/11(72%); Gaps
0/11(0%)

Query  491  QNKFDYDMFFR  501
            Q+K+DY  F R
Sbjct  116  QSKYDYPQFNR  126
```

Alignment statistics for match #3

```
Score 16.9; E-value 3.0; Identities 11/34(32%); Positives 20/34(58%); Gaps
2/34(5%)

Query  5    TAKRSLDVVSPGSLAEFEGSKSR-HDEIENEHRR  37
            T K  LD +SP +   +  S +R  DE+ ++H++
Sbjct  77   TTKHLLDNISP-TFKMYTDSNNRNFDEVNHQHQQ  109
```

Appendix A-5: MAA biosynthetic pathway genes detected in fungi

Fungal species with homology to Keap1, conserved sequences of Keap1 (Kelch1 – Kelch 6 and BTB domain) and with homology to Nrf2 and its conserved sequences Neh1 – Neh6 were analysed for genes encoding mycosporine-like amino acid (MAA) biosynthesis. Genomes were data-mined for pentose-phosphate and shikimate pathways of MAA biosynthesis (Fig A-5.1). Pentose phosphate route of MAA biosynthesis was determined via presence of homology to enzymes Ava_3855 (NRPS-like synthetase), Ava_3856 (ATP-grasp amino acid ligase), Ava_3857 (O-methyltransferase) and Ava_3858 (2-epi-5-epi-valiolone synthase) (599), and the shikimate pathway was determined according to presence of homology to DHQS (3-dehydroquinate synthase) and DAHPS (3-deoxy-d-arabino-heptulosonate 7-phosphate synthase) (600).

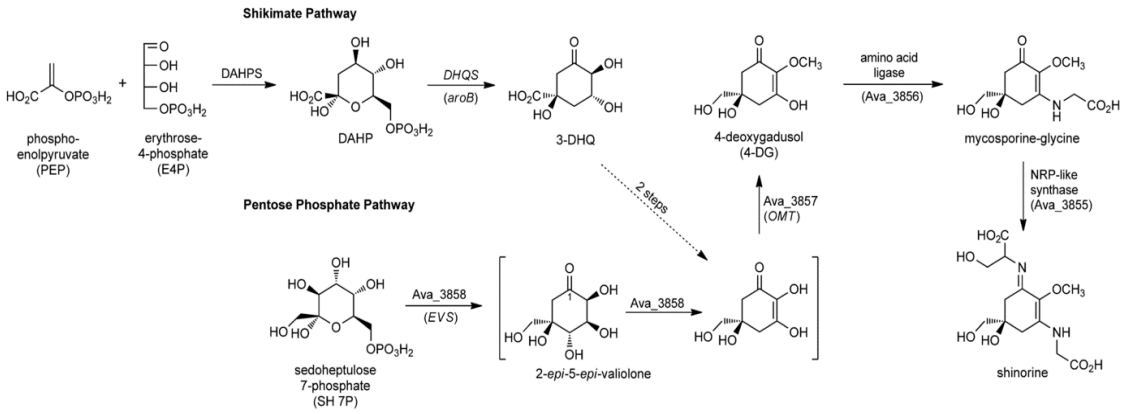


Fig A-5.1. Presumed pathways of MAA biosynthesis, modified from (600)**.**

Results of data mining of fungal genomes for MAA biosynthesis pathways are listed in Table A-5.1.

Table A-5.1. Fungal species with homology to Keap1, Nrf2 and biosynthesis of Mycosporine-like amino acids (MAAs).

Table lists fungal taxa in which homology to Nrf2, Keap1 and at least some enzymes involved in biosynthesis of MAAs could be detected. Nrf2 prediction scores represent the sum of numbers of detected homologs for Nrf2 and domain Neh1 – Neh6 conserved sequences. The Keap1 prediction scores represent the sum of Keap1 sequences detected and the number of conserved Kelch1 – Kelch6 and BTB domains. Species listed in Results (marked in blue) were selected based on Keap1 and Nrf2 prediction scores as well as presence of homology of all tested enzymes involved in MAA biosynthesis. Scores marked with asterisk indicate lack of full-length homolog.

Fungal species	Nrf2 Score	Keap Score	Shikimate pathway enzymes	Pentose-phosphate pathway enzymes
B. bassiana	7	43	2 / 2	4 / 4
C. gloeosporioides	8	79	2 / 2	4 / 4
C. graminicola	15	73	2 / 2	4 / 4
C. higginsianum	14	54	2 / 2	4 / 4
C. militaris	8	46	2 / 2	4 / 4
F. graminearum	10	66	2 / 2	4 / 4
F. pseudograminearum	10	60	2 / 2	4 / 4
M. acridum	13	38	2 / 2	4 / 4
M. oryzae	15	79	2 / 2	4 / 4
V. dahliae	10	40	2 / 2	4 / 4
C. albicans	3	8	1 / 2	2 / 4
C. apollinis	7*	21	2 / 2	4 / 4
C. dubliniensis	10	16	1 / 2	2 / 4
C. globosum	10*	60	1 / 2	4 / 4
C. tenuis	5	5	1 / 2	4 / 4
C. thermophilum	8	22	1 / 2	4 / 4
E. lata	8*	58	1 / 2	4 / 4
K. africana	8*	12	1 / 2	2 / 4
L. elongisporus	6	14	1 / 2	3 / 4
M. anisopliae	7	35	1 / 2	4 / 4
N. crassa	11	71	1 / 2	4 / 4
N. tetrasperma	5	35	1 / 2	4 / 4
P. omphalodes	3*	26	2 / 2	4 / 4
S. cerevisiae	3*	4	2 / 2	2 / 4
S. hirsutum	2*	62	2 / 2	4 / 4
S. macrospora	9	74	1 / 2	4 / 4
S. schenckii	7	26	1 / 2	4 / 4
T. blattae	6	12	1 / 2	1 / 4
T. reesei	5*	25	1 / 2	4 / 4
V. alfalfae	10*	66	2 / 2	2 / 4

Appendix A-6: Keap1 and Nrf2 homologs used for phylogenetic analysis

Appendix A-6 is included in electronic format on a supplied supplementary data files disk and is available online at http://www.sciencedirect.com/science/article/pii/S089158491500283X.

Appendix A-7: Virtual screen results

Table A-7.1. Virtual screening results.

Listed compounds are deemed "viable" according to their docking score, number of potential hydrogen bonds within the Keap1-Nrf2 binding pocket and by manual inspection of the docking profile. The table is sorted according to the biological function of docked ligands.

Compound	Score	Structure Assignment	Biological function
ZINC49048037	0.75	AGN-PC-07CJ71	acetylcholinesterase inhibitor (361)
ZINC15120547	-2.66	Crassinervic acid	antifungal (362)
ZINC00622123	0.77	Griseofulvin	antifungal (363)
ZINC13411177	0.02	similar to Strictifolione	antifungal (364)
ZINC14447808	1.76	AGN-PC-077JEH	antifungal (365)
ZINC40973915	-9.01	similar to Ixoside	antioxidant (366)
ZINC31157290	-2.60	Secoxyloganin	antioxidant (367)
ZINC05998957	-2.17	Lirioresinol A	antioxidant (368)
ZINC15119278	-1.04	similar to Yatein	antioxidant (369)
ZINC00898006	-0.15	Rubrofusarin	antioxidant (370)
ZINC02563652	-0.04	Alloisoimperatorin	antioxidant (371)
ZINC01580260	0.23	Cleomiscosin A	antioxidant (372)
ZINC69482380	1.63	similar to Maclurin	antioxidant (373)
ZINC06037073	-0.97	similar to Emodin	cytotoxic, anti-cancer (374)
ZINC84154280	-2.54	Geranyloxy-p-benzoic Acid	farnesoid X receptor agonist (375)
ZINC26490614	-2.69	Procyanidin B2	Nrf2 activator (376)
ZINC30726399	-9.93	Betanidin	Nrf2 activator (377)
ZINC69482045	-6.62	similar to Ursoloic acid	Nrf2 activator (378)
ZINC69481913	-6.40	similar to Ursoloic acid	Nrf2 activator (378)
ZINC17263588	-6.17	Chlorogenic acid	Nrf2 activator (379)
ZINC84154032	-5.75	similar to Morroniside	Nrf2 activator (380)
ZINC84153764	-4.32	similar to Morroniside	Nrf2 activator (380)
ZINC04102166	-4.28	Geniposidic acid	Nrf2 activator (381)
ZINC01714287	-3.40	Piperine	Nrf2 activator (382)
ZINC03870412	-3.06	Epigallocatechin gallate (EGCG)	Nrf2 activator (383)
ZINC00073693	-2.12	Pinocembrin	Nrf2 activator (384)

ZINC12428433	-1.84	Butein	Nrf2 activator (385)
ZINC71316232	-1.69	similar to Chlorogenic acid	Nrf2 activator (379)
ZINC01531693	-1.57	similar to Piperine	Nrf2 activator (382)
ZINC03872070	-1.52	Chrysine	Nrf2 activator (386)
ZINC00897734	-1.50	similar to Quercetin	Nrf2 activator (387)
ZINC00156701	-1.41	Naringenin	Nrf2 activator (388)
ZINC00113309	1.69	Fraxetin	Nrf2 activator (389)
ZINC01561070	-0.11	similar to Quercetin	Nrf2 activator (387)
ZINC14728348	0.14	similar to Quercetin	Nrf2 activator (387)
ZINC05733652	-1.36	Diosmetin	potential Nrf2 activator, antioxidant (390)
ZINC33832113	-1.73	similar to Phlorizin	potential Nrf2 activator (391)
ZINC69482290	-3.37	similar to Glucoerucin	potential Nrf2 activator (392)
ZINC05733537	-0.86	Ermanin, similar to Quercetin (Nrf2 activator)	potential Nrf2 activator (387)
ZINC84153966	-3.86	similar to Acetoside	potential Nrf2 activator (393)
ZINC13108875	-2.42	similar to Burchellin	potential pesticide (394)
mycosporine glycine-valine	0.24	Mycosporine-like amino acid	UV-protectant, antioxidant (395)
mycosporine glycine	4.21	Mycosporine-like amino acid	UV-protectant, antioxidant (395)
Porphyra 334	0.58	Mycosporine-like amino acid	UV-protectant, antioxidant (395)
ZINC15252691	-5.23	Gaudichaudianic acid	trypancide (396)
ZINC36403425	-6.02	no annotation available	unknown
ZINC69482210	-5.93	no annotation available	unknown
ZINC69482050	-5.86	no annotation available	unknown
ZINC69482369	-5.49	no annotation available	unknown
ZINC69482017	-4.92	no annotation available	unknown
ZINC69481856	-4.71	no annotation available	unknown
ZINC84153792	-4.66	no annotation available	unknown
ZINC84154198	-4.61	no annotation available	unknown
ZINC84154010	-4.16	no annotation available	unknown
ZINC84154210	-3.80	no annotation available	unknown
ZINC84154686	-3.33	no annotation available	unknown
ZINC84154382	-3.29	no annotation available	unknown
ZINC84153978	-2.93	no annotation available	unknown
ZINC69481928	-2.91	no annotation available	unknown
ZINC84153941	-2.54	no annotation available	unknown
ZINC69482062	-1.89	no annotation available	unknown

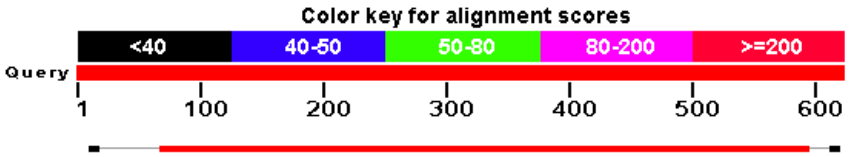
ZINC69482080	-1.51	no annotation available	unknown
ZINC69482112	-1.48	no annotation available	unknown
ZINC84154336	-1.46	no annotation available	unknown
ZINC84154341	-1.37	no annotation available	unknown
ZINC00174016	-1.28	no annotation available	unknown
ZINC69482027	-1.07	no annotation available	unknown
ZINC84154296	-1.00	no annotation available	unknown
ZINC84154672	-0.88	no annotation available	unknown
ZINC84154479	-0.87	no annotation available	unknown
ZINC14647433	-0.31	no annotation available	unknown
ZINC19204246	0.53	no annotation available	unknown
ZINC14445187	0.77	hydroxycanthin	unknown
ZINC69481958	1.05	no annotation available	unknown
ZINC15119468	1.21	no annotation available	unknown

Appendix A-8: *Caenorhabditis elegans* Keap1 homology analysis

Alignment of Human Keap1 and C. elegans protein SKN1

No significant similarity found (within BLAST e-value 100.0)

Alignment of Human Keap1 and *C. elegans* Keap1-like sequence R12E2.1 (SKN1)

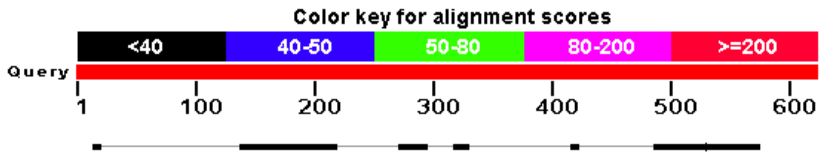


Statistics: Score: 574 bits; Coverage: 87%; BLAST E-value: 5.0e-116

```
Query  52   FSYTLEDHTKQAFGIMNELRLSQQLCDVTLQVKYQDAPAAQFMAHKVVLASSSPVFKAMF  111
            ++ + ++ K+A  +M  +R    L DV L+VK +      F AHKVVL+++SP FKAMF
Sbjct  65   MTFCMSNYAKEALKMMYMMRSHGMLTDVVLEVKKE-----LFPAHKVVLSAASPYFKAMF  119

Query  112  TNGLREQGMEVVSIEGIHPKVMERLIEFAYTASISMGEKCVLHVMNGAVMYQIDSVVRAC  171
            T GL+E  M  V ++G+ P  M R++ F YT  I + E  V  ++  A M+Q+ +V+ AC
Sbjct  120  TGGLKESEMSRVQLQGVCPTAMSRILYFMYTGQIRVTEVTVCQLLPAATMFQVPNVIDAC  179

Query  172  SDFLVQQLDPSNAIGIANFAEQIGCVELHQRAREYIYMHFGEVAKQEEFFNLSHCQLVTL  231
              FL +QLDP+NAIGIA+FAEQ GCVEL ++A  +I  +F +V ++EEF  LS  QL+ L
Sbjct  180  CAFLERQLDPTNAIGIAHFAEQHGCVELQKKANVFIERNFTQVCQEEEFLQLSAYQLIAL  239

Query  232  ISRDDLNVRCESEVFHACINWVKYDCEQRRFYVQALLRAVRCHSLTPNFLQMQLQKCEIL  291
            I RD+LNV+ E EV++A + WVKYD + R   ++ +L AVRC  LTPNFL+ Q++ C++L
Sbjct  240  IRRDELNVQEEREVYNAVLKWVKYDEDNRHCKMEHILGAVRCQFLTPNFLKEQMKNCDVL  299
```

```
Query  292  QSDSRCKDYLVKIFEELTLHKPTQVMPCRAPKVGRLIYTAGGYFRQSLSYLEAYNPSDGT  351
            +    C++YL KIF++LTLHK   V   R P   R+I+ AGG+FR SL  LEAYN  D T
Sbjct  300  RKVPACREYLAKIFKDLTLHKCPGVKE-RTPNTTRMIFVAGGFFRHSLDILEAYNVDDMT  358

Query  352  WLRLADLQVPRSGLAGCVVGGLLYAVGGRNNSPDGNTDSSALDCYNPMTNQWSPCAPMSV  411
            W  LA+L++PRSGL    + G  YAVGGRNN+   + DS  +D Y+ +T  W PCAPMSV
Sbjct  359  WTTLANLRIPRSGLGAAFLKGKFYAVGGRNNNIGSSYDSDWVDRYSAVTETWRPCAPMSV  418

Query  412  PRNRIGVGVIDGHIYAVGGSHGCIHHNSVERYEPERDEWHLVAPMLTRRIGVGVAVLNRL  471
            PR+R+GV V+D  +YAVGGS G  +HN+VE Y+P+ D W LV PM  +R+GVGV V+NRL
Sbjct  419  PRHRVGVAVMDELMYAVGGSAGMEYHNTVEYYDPDLDRWTLVQPMHAKRLGVGVVVVNRL  478

Query  472  LYAVGGFDGTNRLNSAECYYPERNEWRMITAMNTIRSGAGVCVLHNCIYAAGGYDGQDQL  531
            LYA+GGFDG  RL S ECY+PE NEW  +  + T RSGAGV  ++  IY  GG+DG  QL
Sbjct  479  LYAIGGFDGNERLASVECYHPENNEWSFLPPLQTGRSGAGVAAINQYIYVVGGFDGTRQL  538

Query  532  NSVERYDVETETWTFVAPMKHRRSALGITVHQGRIYVLGGYDGHTFLDSVECYDPDTDTW  591
             +VERYD E +TW  VAP++  RSAL +T    ++Y +GG+DG+ FL  VE YDP T+TW
Sbjct  539  ATVERYDTENDTWDMVAPIQIARSALSLTPLDEKLYAIGGFDGNNFLSIVEVYDPRTNTW  598

Query  592  SEVTRMTSGRSGVGVAVTMEP  612
            +  T + SGRSG   AV  +P
Sbjct  599  TTGTPLKSGRSGHASAVIYQP  619
```

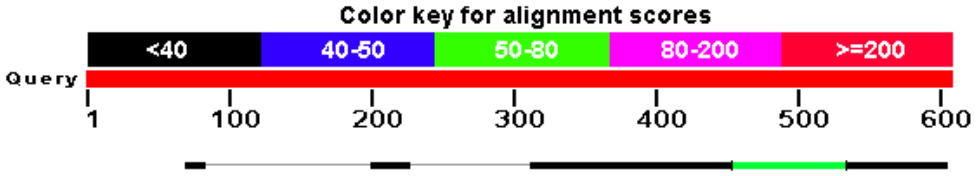
Alignment of Human Keap1 and *C. elegans* protein WDR-23



Statistics: Score: 16.5; Coverage 36%; BLAST E-value: 3.8

```
Query  319  CRAPKVGRLIYTAG  332
            C A     LIY+AG
Sbjct  318  CFADLGSNLIYSAG  331

Query  488  EC-YYPERNE-----WRMITAMNTI---RSGAGVCVLHNCIYAAGGYDGQDQL  531
            EC ++P  NE     W  +T + T    + G      H  I   G  D  D+L
Sbjct  489  ECDWHPTENEIVSSAWDGVTTVWTWDERQEGVIAPYDHPNISQFGDEDSCDEL  541
```

Alignment of Human β-transducin repeat-containing protein (β-TrCP) and *C. elegans* protein WDR-23



Statistics: Score 62.4; Coverage: 55%; BLAST e-value: 2.0e-14

```
Query  339  LTGHTG----SVLCLQYDE--RVIITGSSDSTVRVWDVNTGEMLNTLIHHCE----AVLH  388
            L G  G     +V C+++ +     I+ G+S  ++ V+DV      + T+++  E     +V
Sbjct  260  LNGEPGRDHCAVFCVKFSDSSEQIVCGTSQYSIHVFDVEQRRRIRTIVNAHEDDVNSVCF  319

Query  389  LRFNNGMMVTCSKDRSIAVWDMA--SPTDITLRRVLVGHRAAVNVVDF--DDKYIVSASG  444
                + ++ +    D  + VWD    S  D+    V  GHR  V  VD   D++Y++S S
Sbjct  320  ADLGSNLIYSAGDDGLVKVWDKRAWSDGDVEPVGVFAGHRDGVTHVDSRQDERYLLSNSK  379

Query  445  DRTIKVW------NTSTCEFVR------------------------------TLNGH---  465
            D+TIKVW      N S  E  R                               TL GH
Sbjct  380  DQTIKVWDLRKFSNMSGVEATRACVQSQHWDYRWQPAPPGLCQPVAGDTSVMTLRGHSVL  439

Query  466  ----KRGIACLQYRDRLVVSGSSDNTIRLWDIECGACLRVLEGHEELVRCIRFD--NKRI  519
                +   +      R + +G +   + ++DI  G   R L+GH  +VR   +      I
Sbjct  440  HTLVRANFSPESTGRRYIYTGCARGEVVVYDIMSGTVSRRLKGHTAVVRECDWHPTENEI  499

Query  520  VSGAYDGKIKVW  531
            VS A+DG   VW
Sbjct  500  VSSAWDGVTTVW  511
```

Appendix B

Appendix B presents a large number of protein sequences used in phylogenetic analysis and is unsuitable for paper-format. It is included in electronic format on a supplied supplementary data files disk as Appendix B.

Appendix is also available online at http://www.nature.com/articles/srep27740#supplementary-information.

Appendix C

Appendix C-1: UV spectrum of twin Philips Ultraviolet-B TL 20W/12RS lamps

The appendix C-1 presents the UV spectrum (Fig C-1.1) of lamps used to irradiate yeast cultures in the experiments described in the Chapter 5, and the detailed breakdown of the UV output of lamps, calculated for different bands of UV spectrum (Table C-1.1).

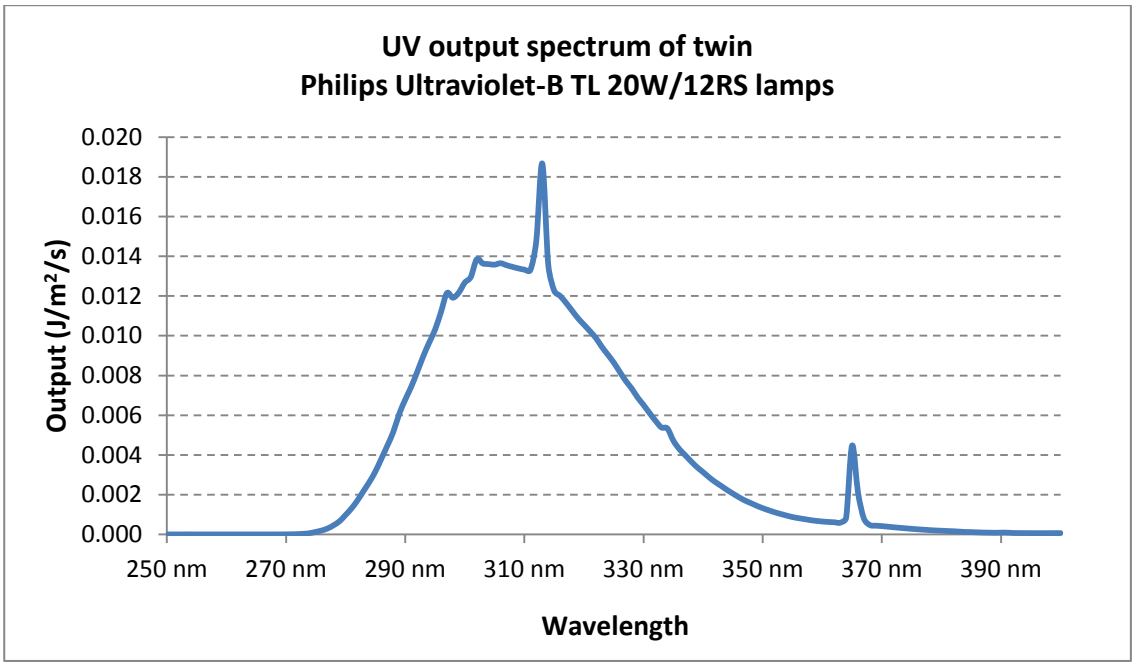


Fig C-1.1: Irradiance spectrum of twin Philips Ultraviolet-B TL 20W/12RS lamps

The figure displays a UV emission spectrum of twin Philips Ultraviolet-B TL 20W/12RS lamps used to irradiate yeast samples. The UV output of lamps was measured at 10 cm distance from the lamps, using Bentham (double grating) DM150BC spectroradiometer with 2 400 g/mm grating blazed at 250 nm, IS4C integration sphere diffuser, DH3 (bi) photomultiplier and large aperture.

Table C-1.1. UV output of twin Philips Ultraviolet-B TL 20W/12RS lamps.

Table lists the calculations of UV output of twin Philips Ultraviolet-B TL 20W/12RS lamps for different bands of UV spectrum. The main bands of UV spectrum (UV-A, UV-B and UV-C) are underlined for clarity. The output of twin lamps, corresponding wavelengths and the proportion of total UV output are listed for each band of UV spectrum.

UV band	Wavelength (nm)	UV output ($J/m^2/s$)	Proportion of total UV
UV-A	321-400	1.75	29.9%
(UV-A > 315)	316-400	2.31	39.6%
UV-A I	341-400	0.45	7.7%
UV-A II	321-340	1.30	22.2%
(UV A II > 315)	316-340	1.86	31.9%
UV-B	281-320	4.06	69.6%
Non-solar UV-B	281-295	0.82	14.0%
UV-C	251-280	0.03	0.5%
Total UV	250-400	5.84	100.0%

Appendix C-2: Quantified proteome of *Sporobolomyces* yeast LEV-2

Appendix C-2 presents a list of all proteins identified and quantified in MudPIT analysis of UV-tolerant *Sporobolomyces* yeast LEV-2, and is unsuitable for paper-format. It is included in electronic format on a supplied supplementary data files disk as Appendix C.

Bibliography

1. Sies H. Oxidative stress: introductory remarks. In: Oxidative stress. Academic Press London; 1985. p. 1–8.

2. Commoner B, Townsend J, Pake GE. Free radicals in biological materials. Nature. 1954;174(4432):689–91.

3. Mccords JM, Fridovich I. Superoxide dismutase: an enzymic function for erythrocuprein (hemocuprein). J Biol Chem. 1969;244(22):6049–65.

4. Harman D. Aging: a theory based on free radical and radiation chemistry. J Gerontol. 1956;11(3):298–300.

5. Sies H, Pruchniak MP, Araźna M, Demkc U. Biochemistry of oxidative stress. Angew Chemie Int Ed English. 1986;25(1):1058–71.

6. Halliwell B, Gutteridge JMC. Oxygen toxicity, oxygen radicals, transition metals and disease. Biochem J. 1984;219:1–14.

7. Home - PubMed - NCBI [Internet]. [cited 2017 Jan 18]. Available from: https://www.ncbi.nlm.nih.gov/pubmed

8. Halliwell B, Gutteridge J. Lipid peroxidation, oxygen radicals, cell damage and antioxidant therapy. Lancet. 1984;4:1396–7.

9. Gutteridge JMC, Halliwell B. Free radicals in disease processes: a compilation of cause and consequence. Free Radic Res. 1993;19(3):141–58.

10. Suthanthiran M, Solomon SD, Williams PS, Rubin AL, Novogrodsky A, Stenzel KH. Hydroxyl radical scavengers inhibit human natural killer cell activity. Vol. 307, Nature. 1984. p. 276–8.

11. Lebovitz RM, Zhang H, Vogel H, Cartwright J, Dionne L, Lu N, et al. Neurodegeneration, myocardial injury, and perinatal death in mitochondrial superoxide dismutase-deficient mice. Proc Natl Acad Sci U S A. 1996;93(18):9782–7.

12. Kamata H, Hirata H. Redox regulation of cellular signalling. Cell Signal. 1999;11(1):1–14.

13. Sies H. Hydrogen peroxide as a central redox signaling molecule in physiological oxidative stress: oxidative eustress. Redox Biol. 2017;11:613–9.

14. Forman HJ, Maiorino M, Ursini F. Signaling functions of reactive oxygen species. Biochemistry. 2010;49(5):835–42.

15. Ristow M, Zarse K, Oberbach A, Klöting N, Birringer M, Kiehntopf M, et al. Antioxidants prevent health-promoting effects of physical exercise in humans. Proc Natl Acad Sci U S A. 2009;106(21):8665–70.

16. Gutteridge JMC, Halliwell B. Antioxidants: molecules, medicines, and myths. Biochem Biophys Res Commun. 2010;393(4):561–4.

17. Forman HJ, Augusto O, Brigelius-Flohe R, Dennery PA, Kalyanaraman B, Ischiropoulos H, et al. Even free radicals should follow some rules: a guide to free radical research terminology and methodology. Free Radic Biol Med. 2015;78:233–5.

18. Halliwell B, Gutteridge JMC. Role of free radicals and catalytic metal ions in human disease: An overview. Methods Enzymol. 1990;186:1–85.

19. Finkel T, Holbrook NJ. Oxidants, oxidative stress and the biology of ageing. Nature. 2000;408(6809):239–47.

20. Madamanchi NR, Runge MS. Redox signaling in cardiovascular health and disease. Free Radic Biol Med. 2013;61:473–501.

21. Feig DI, Reid TM, Loeb LA. Reactive oxygen species in tumorigenesis. Cancer Res. 1994;54:1890s–1894s.

22. Chen Y, Gibson SB. Is mitochondrial generation of reactive oxygen species a trigger for autophagy? Autophagy. 2008;4(2):246–8.

23. Sies H, Jones D. Oxidative stress. In: Encyclopedia of Stress. 2010. p. 45–8.

24. Sarsour EH, Kalen AL, Goswami PC. Manganese superoxide dismutase regulates a redox cycle within the cell cycle. Antioxid Redox Signal. 2014;20(10):1618–27.

25. Jones DP. Redefining oxidative stress. Antioxid Redox Signal. 2006;8(9,10):1865–79.

26. Halliwell B. How to characterize an antioxidant: an update. Biochem Soc Symp. 1995;61:73–101.

27. Boveris BA, Chance B. The mitochondrial generation of hydrogen peroxide. Biochem J. 1973;134:707–16.

28. Murphy MP. How mitochondria produce reactive oxygen species. Biochem J. 2009;417:1–13.

29. Holmström KM, Finkel T. Cellular mechanisms and physiological consequences of redox-dependent signalling. Nat Rev Mol Cell Biol. 2014;15(6):411–21.

30. Circu ML, Aw TY. Reactive oxygen species, cellular redox systems, and apoptosis. Free Radic Biol Med. 2010;48(6):749–62.

31. Winterbourn CC. Reconciling the chemistry and biology of reactive oxygen species. Nat Chem Biol. 2008;4(5):278–86.

32. Krumova K, Cosa G. Chapter I: Overview of reactive oxygen species. In: Nonell S, Flors C, editors. Singlet Oxygen: Applications in Biosciences and Nanosciences, Volume 1. Royal Society of Chemistry; 2016. p. 1–21.

33. Klaunig JE, Kamendulis LM. The role of oxidative stress in carcinogenesis. Annu Rev Pharmacol Toxicol. 2004;44:239–67.

34. Diffey BL. Solar ultraviolet radiation effects on biological systems. Phys Med Biol. 1991;36(3):299–328.

35. The International Organization for Standardization. ISO 21348 Definitions of Solar Irradiance Spectral Categories. Environment. 2007;(section 5):6–7.

36. Madronich S. The atmosphere and UV-B radiation at ground level. In: Young AR, Bjorn LO, Moan J, Nultsch W, editors. Environmental UV Photobiology. Boston, MA: Springer US; 1993. p. 1–39.

37. Yoon JH, Lee CS, O'Connor TR, Yasui A, Pfeifer GP. The DNA damage spectrum produced by simulated sunlight. J Mol Biol. 2000;299(3):681–93.

38. Davies MJ, Truscott RJ. Photo-oxidation of proteins and its role in cataractogenesis. J Photochem Photobiol. 2001;63(1–3):114–25.

39. Heck DE, Vetrano AM, Mariano TM, Laskin JD. UVB light stimulates production of reactive oxygen species: Unexpected role for catalase. J Biol Chem. 2003;278(25):22432–6.

40. Svobodová AR, Galandáková A, Sianská J, Doležal D, Lichnovská R, Ulrichová J, et al. DNA damage after acute exposure of mice skin to physiological doses of UVB and UVA light. Arch Dermatol Res. 2012;304(5):407–12.

41. Schwarz T. Mechanisms of UV-induced immunosuppression. Keio J Med. 2005;54(4):165–71.

42. Findlay G. Ultra-violet light and skin cancer. Lancet. 1928;1070–3.

43. Brash DE, Rudolph J a, Simon J a, Lin A, McKenna GJ, Baden HP, et al. A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. Proc Natl Acad Sci U S A. 1991;88(22):10124–8.

44. Moran DJ, Hollows FC. Pterygium and ultraviolet radiation: a positive correlation. Br J Ophthalmol. 1984;68(5):343–6.

45. Petersen M. Ultraviolet A irradiation stimulates collagenase production in cultured human fibroblasts. J Invest Dermatol. 1992;99(4):440–4.

46. Wang SQ, Setlow R, Berwick M, Polsky D, Marghoob AA, Kopf AW, et al. Ultraviolet A and melanoma: a review. J Am Acad Dermatol. 2001;44(5):837–46.

47. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the *BRAF* gene in human cancer. Nature. 2002;417(6892):949–54.

48. Finkel T. Signal transduction by reactive oxygen species. J Cell Biol. 2011;194(1):7–15.

49. Breimer LH. Molecular mechanisms of oxygen radical carcinogenesis and mutagenesis: the role of DNA base damage. Mol Carcinog. 1990;3(4):188–97.

50. Ma Q. Transcriptional responses to oxidative stress: pathological and toxicological implications. Pharmacol Ther. 2010;125(3):376–93.

51. Sies H. Strategies of antioxidant defense. Eur J Biochem. 1993;215(2):213–9.

52. Sies H. Oxidative stress: oxidants and antioxidants. Exp Physiol. 1997;82(1):291–5.

53. Halliwell B, Gutteridge JMC. Free Radicals in Biology and Medicine. Fifth Edit. Halliwell B, Gutteridge JMC, editors. Oxford: Oxford University Press; 2015.

54. Birben E, Sahiner UM, Sackesen C, Erzurum S, Kalayci O. Oxidative stress and antioxidant defense. WAO J. 2012;5(1):9–19.

55. Halliwell B, Chirico S. Lipid peroxidation: its mechanism, measurement, and significance. Am J Clin Nutr. 1993;57:715S–25S.

56. Girotti A. Mechanisms of lipid peroxidation. J Free Radic Biol Med. 1985;1(1):87–95.

57. Dalle-Donne I, Rossi R, Colombo R, Giustarini D, Milzani A. Biomarkers of oxidative damage in human disease. Clin Chem. 2006;52(4):601–23.

58. Berlett BS, Stadtman ER. Protein oxidation in aging, disease, and oxidative stress. J Biol Chem. 1997;272(33):20313–6.

59. Jones DP. Radical-free biology of oxidative stress. Am J Physiol Cell Physiol. 2008;295(4):C849-68.

60. Zhang R, Al-Lamki R, Bai L, Streb JW, Miano JM, Bradley J, et al. Thioredoxin-2 inhibits mitochondria-located ASK1-mediated apoptosis in a JNK-independent manner. Circ Res. 2004;94(11):1483–91.

61. Jones DP, Sies H. The redox code. Antioxid Redox Signal. 2015;23(9):734–46.

62. Kump LR. The rise of atmospheric oxygen. Nature. 2008;451(7176):277–8.

63. Fenchel T. Anaerobic Eukaryotes. In: Altenbach A, Bernhard JM, Seckbach J, editors. Anoxia. Springer Netherlands; 2012. p. 3–16.

64. Shick JM, Dunlap WC. Mycosporine-like amino acids and related gadusols: biosynthesis, accumulation, and UV-protective functions in aquatic organisms. Annu Rev Physiol. 2002;64(1):223–62.

65. Sancar GB, Sancar A. Structure and function of DNA photolyases. FEBS Lett. 1987;12(July):259–61.

66. Meneghini R. Iron homeostasis, oxidative stress, and DNA damage. Free Radic Biol Med. 1997;23(5):783–92.

67. Kensler TTW, Wakabayashi N, Biswal S. Cell survival responses to environmental stresses via the Keap1-Nrf2-ARE pathway. Annu Rev Pharmacol Toxicol. 2007;47(1):89–116.

68. Nordberg J, Arner E. Reactive oxygen species, antioxidants, and the mammalian thioredoxin system. Free Radic Biol Med. 2001;31(11):1287–312.

69. Barnham KJ, Masters CL, Bush AI. Neurodegenerative diseases and oxidative stress. Nat Rev Drug Discov. 2004;3(3):205–14.

70. Madamanchi NR, Vendrov A, Runge MS. Oxidative stress and vascular disease. Arterioscler Thromb Vasc Biol. 2005;25(1):29–38.

71. Giacco F, Brownlee M. Oxidative stress and diabetic complications. Circ Res. 2010;107(9):1058–70.

72. Hussain SP, Hofseth LJ, Harris CC. Radical causes of cancer. Nat Rev Cancer. 2003;3(4):276–85.

73. Zhang M, An C, Gao Y, Leak RK, Chen J, Zhang F. Emerging roles of Nrf2 and phase II antioxidant enzymes in neuroprotection. Prog Neurobiol. 2013;100(1):30–47.

74. Iizuka T, Ishii Y, Itoh K, Kiwamoto T, Kimura T, Matsuno Y, et al. Nrf2-deficient mice are highly susceptible to cigarette smoke-induced emphysema. Genes to Cells. 2005;10(12):1113–25.

75. Khor TO, Huang MT, Prawan A, Liu Y, Hao X, Yu S, et al. Increased susceptibility of Nrf2 knockout mice to colitis-associated colorectal cancer. Cancer Prev Res. 2008;1(3):187–91.

76. Viña J, Borrás C, Miquel J. Theories of ageing. IUBMB Life. 2007;59(4–5):249–54.

77. Grignolio A, Franceschi C. History of Research into ageing/senescence. eLS. 2012;(June):1–9.

78. Jin K. Modern biological theories of aging. Aging Dis. 2010;1(2):72–4.

79. Barja G. Updating the mitochondrial free radical theory of aging: an integrated view, key aspects, and confounding concepts. Antioxid Redox Signal. 2013;19(12):1420–45.

80. Kowaltowski AJ, de Souza-Pinto NC, Castilho RF, Vercesi AE. Mitochondria and reactive

oxygen species. Free Radic Biol Med. 2009;47(4):333–43.

81. Harman D. Free radical theory of aging: consequences of mitochondrial aging. Age (Omaha). 1983;6:86–94.

82. Buettner GR. Moving free radical and redox biology ahead in the next decade(s). Free Radic Biol Med. 2015;78:236–8.

83. Sies H. Oxidative stress: A concept in redox biology and medicine. Redox Biol. 2015;4:180–3.

84. Navarro A, Gomez C, López-Cepero JM, Boveris A. Beneficial effects of moderate exercise on mice aging: survival, behavior, oxidative stress, and mitochondrial electron transfer. Am J Physiol Regul Integr Comp Physiol. 2004;286(3):R505-11.

85. Bjelakovic G, Nikolova D, Gluud LL, Simonetti RG, Gluud C. Antioxidant supplements for prevention of mortality in healthy participants and patients with various diseases. Sao Paulo Med J. 2015;133(2):164.

86. Andersen JK. Oxidative stress in neurodegeneration: cause or consequence? Nat Med. 2004;10(July):S18-25.

87. Liochev SI. Reactive oxygen species and the free radical theory of aging. Free Radic Biol Med. 2013;60:1–4.

88. Salmon AB, Richardson A, Pérez VI. Update on the oxidative stress theory of aging: does oxidative stress play a role in aging or healthy aging? Free Radic Biol Med. 2010;48(5):642–55.

89. Lushchak VI. Free radicals, reactive oxygen species, oxidative stress and its classification. Chem Biol Interact. 2014;224:164–75.

90. Forman HJ, Davies KJA, Ursini F. How do nutritional antioxidants really work: nucleophilic tone and para-hormesis versus free radical scavenging in vivo. Free Radic Biol Med. 2014;66:24–35.

91. Bjelakovic G, Nikolova D, Gluud C. Antioxidant supplements and mortality. Curr Opin Clin Nutr Metab Care. 2014;17(1):40–4.

92. Benedict AL, Knatko E V., Dinkova-Kostova AT. The indirect antioxidant sulforaphane protects against thiopurine-mediated photooxidative stress. Carcinogenesis. 2012;33(12):2457–66.

93. Kode A, Rajendrasozhan S, Caito S, Yang S-R, Megson IL, Rahman I. Resveratrol induces glutathione synthesis by activation of Nrf2 and protects against cigarette smoke-mediated oxidative stress in human lung epithelial cells. Am J Physiol Lung Cell Mol Physiol. 2008;294(3):L478-88.

94. Halliwell B, Rafter J, Jenner A. Health promotion by flavonoids, tocopherols, tocotrienols, and other phenols: direct or indirect effects? Antioxidant or not? Am J Clin Nutr. 2005;81(1 Suppl):268–76.

95. Dinkova-Kostova AT, Talalay P. Direct and indirect antioxidant properties of inducers of cytoprotective proteins. Mol Nutr Food Res. 2008;52(Suppl.1):S128–38.

96. Hayes JD, Dinkova-Kostova AT. The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. Trends Biochem Sci. 2014;39(4):199–218.

97. Halliwell B. The antioxidant paradox. Lancet. 2000;355:1179–80.

98. Halliwell B. Free radicals and antioxidants - quo vadis? Trends Pharmacol Sci. 2011;32(3):125–

30.

99. Halliwell B. The wanderings of a free radical. Free Radic Biol Med. 2009;46(5):531–42.

100. Juránek I, Bezek S. Controversy of free radical hypothesis: reactive oxygen species-cause or consequence of tissue injury? Gen Physiol Biophys. 2005;24:263–78.

101. Maramag C, Menon M, Balaji KC, Reddy PG, Laxmanan S. Effect of vitamin C on prostate cancer cells in vitro: effect on cell number, viability, and DNA synthesis. Prostate. 1997;32(3):188–95.

102. Pauling L. Effect of ascorbic mammary tumors acid on incidence and UV-light-induced of spontaneous skin tumors. Am J Clin Nutr. 1991;54:1252S–5S.

103. Yeom C-H, Lee G, Park J-H, Yu J, Park S, Yi S-Y, et al. High dose concentration administration of ascorbic acid inhibits tumor growth in BALB/C mice implanted with sarcoma 180 cancer cells via the restriction of angiogenesis. J Transl Med. 2009;7:70.

104. McIntyre BS, Briski KP, Gapor A, Sylvester PW. Antiproliferative and apoptotic effects of tocopherols and tocotrienols on preneoplastic and neoplastic mouse mammary epithelial cells. Proc Soc Exp Biol Med. 2008;224(4):292–301.

105. Beal MF, Oakes D, Shoulson I, Henchcliffe C, Galpern WR, Haas R, et al. A randomized clinical trial of high-dosage coenzyme Q10 in early Parkinson Disease. JAMA Neurol. 2014;71(5):543.

106. Galasko DR, Peskind E, Clark CM, Quinn JF, Ringman JM, Jicha GA, et al. Antioxidants for Alzheimer Disease. Arch Neurol. 2012;69(7):3036–40.

107. Goodman M, Bostick RM, Kucuk O, Jones DP. Clinical trials of antioxidants as cancer prevention agents: past, present, and future. Free Radic Biol Med. 2011;51(5):1068–84.

108. Firuzi O, Miri R, Tavakkoli M, Saso L. Antioxidant therapy: current status and future prospects. Curr Med Chem. 2011;18(25):3871–88.

109. Blumberg JB, Frei B. Why clinical trials of vitamin E and cardiovascular diseases may be fatally flawed. Commentary on "The relationship between dose of vitamin E and suppression of oxidative stress in humans." Free Radic Biol Med. 2007;43(10):1374–6.

110. Steinhubl SR. Why have antioxidants failed in clinical trials? Am J Cardiol. 2008;101(10 Suppl.):14D–19D.

111. Hybertson BM, Gao B, Bose SK, McCord JM. Oxidative stress in health and disease: the therapeutic potential of Nrf2 activation. Mol Aspects Med. 2011;32(4):234–46.

112. Suzuki T, Motohashi H, Yamamoto M. Toward clinical application of the Keap1-Nrf2 pathway. Trends Pharmacol Sci. 2013;34(6):340–6.

113. Keum Y-SS, Choi BY. Molecular and chemical regulation of the Keap1-Nrf2 signaling pathway. Molecules. 2014;19(7):10074–89.

114. Rushmore TH, Pickett CB. Transcriptional regulation of the rat glutathione S-transferase Ya subunit gene. Characterization of a xenobiotic-responsive element controlling inducible expression by phenolic antioxidants. J Biol Chem. 1990;265(24):14648–53.

115. Rushmore TH, Morton MR, Pickett CB. The antioxidant responsive element: Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity. J Biol Chem. 1991;266(18):11632–9.

116. Friling RS, Bensimon A, Tichauer Y, Daniel V. Xenobiotic-inducible expression of murine glutathione S-transferase Ya subunit gene is controlled by an electrophile-responsive element. Proc Natl Acad Sci U S A. 1990;87(16):6258–62.

117. Nerland DE. The antioxidant/electrophile response element motif. Drug Metab Rev. 2007;39(1):235–48.

118. Nioi P, McMahon M, Itoh K, Yamamoto M, Hayes JD. Identification of a novel Nrf2-regulated antioxidant response element (ARE) in the mouse NAD(P)H:quinone oxidoreductase 1 gene: Reassessment of the ARE consensus sequence. Biochem J. 2003;374(Pt 2):337–48.

119. Moi P, Chant K, Asunis I, Cao A, Kant YW, Kan YW. Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the f-globin locus control region. Genetics. 1994;91(21):9926–30.

120. Itoh K, Igarashi K, Hayashi N, Nishizawa M, Yamamoto M. Cloning and characterization of a novel erythroid cell-derived CNC family transcription factor heterodimerizing with the small Maf family proteins. Mol Cell Biol. 1995;15(8):4184–93.

121. Chan K, Lu R, Chang JC, Kan YW. NRF2, a member of the NFE2 family of transcription factors, is not essential for murine erythropoiesis, growth, and development. Proc Natl Acad Sci U S A. 1996;93(24):13943–8.

122. Itoh K, Chiba T, Takahashi S, Ishii T, Igarashi K, Katoh Y, et al. An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. Biochem Biophys Res Commun. 1997;236(2):313–22.

123. Yamazaki H, Katsuoka F, Motohashi H, Engel JD, Yamamoto M. Embryonic lethality and fetal liver apoptosis in mice lacking all three small Maf proteins. Mol Cell Biol. 2012;32(4):808–16.

124. Chanas SA, Jiang Q, McMahon M, McWalter GK, McLellan LI, Elcombe CR, et al. Loss of the Nrf2 transcription factor causes a marked reduction in constitutive and inducible expression of the glutathione S-transferase Gsta1, Gsta2, Gstm1, Gstm2, Gstm3 and Gstm4 genes in the livers of male and female mice. Biochem J. 2002;365(2):405–16.

125. Malhotra D, Portales-Casamar E, Singh A, Srivastava S, Arenillas D, Happel C, et al. Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through chip-seq profiling and network analysis. Nucleic Acids Res. 2010;38(17):5718–34.

126. Yates MS, Tran QT, Dolan PM, Osburn WO, Shin S, McCulloch CC, et al. Genetic versus chemoprotective activation of Nrf2 signaling: overlapping yet distinct gene expression profiles between Keap1 knockout and triterpenoid-treated mice. Carcinogenesis. 2009;30(6):1024–31.

127. Agyeman AS, Chaerkady R, Shaw PG, Davidson NE, Visvanathan K, Pandey A, et al. Transcriptomic and proteomic profiling of KEAP1 disrupted and sulforaphane-treated human breast epithelial cells reveals common expression profiles. Breast Cancer Res Treat. 2012;132(1):175–87.

128. Hirotsu Y, Katsuoka F, Funayama R, Nagashima T, Nishida Y, Nakayama K, et al. Nrf2-MafG heterodimers contribute globally to antioxidant and metabolic networks. Nucleic Acids Res. 2012;40(20):10228–39.

129. Chorley BN, Campbell MR, Wang X, Karaca M, Sambandan D, Bangura F, et al. Identification of novel NRF2-regulated genes by ChiP-Seq: influence on retinoid X receptor alpha. Nucleic Acids Res. 2012;40(15):7416–29.

130. Maher JM, Dieter MZ, Aleksunes LM, Slitt AL, Guo G, Tanaka Y, et al. Oxidative and electrophilic stress induces multidrug resistance-associated protein transporters via the nuclear

factor-E2-related factor-2 transcriptional pathway. Hepatology. 2007;46(5):1597–610.

131. Itoh K, Wakabayashi N, Katoh Y, Ishii T, Igarashi K, Engel JD, et al. Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. Genes Dev. 1999;13(1):76–86.

132. Dhakshinamoorthy S, Jaiswal AK. Functional characterization and role of INrf2 in antioxidant response element-mediated expression and antioxidant induction of NAD(P)H:quinone oxidoreductase1 gene. Oncogene. 2001;20(29):3906–17.

133. Itoh K, Wakabayashi N, Katoh Y, Ishii T, O'Connor T, Yamamoto M. Keap1 regulates both cytoplasmic-nuclear shuttling and degradation of Nrf2 in response to electrophiles. Genes to Cells. 2003;8(4):379–91.

134. Kobayashi A, Kang M, Okawa H, Zenke Y, Chiba T, Igarashi K, et al. Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2. Mol Cell Biol. 2004;24(16):7130–9.

135. Tong KI, Padmanabhan B, Kobayashi A, Shang C, Hirotsu Y, Yokoyama S, et al. Different electrostatic potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response. Mol Cell Biol. 2007;27(21):7511–21.

136. Baird L, Llères D, Swift S, Dinkova-Kostova AT. Regulatory flexibility in the Nrf2-mediated stress response is conferred by conformational cycling of the Keap1-Nrf2 protein complex. Proc Natl Acad Sci U S A. 2013;110(38):15259–64.

137. Baird L, Swift S, Llères D, Dinkova-Kostova AT. Monitoring Keap1-Nrf2 interactions in single live cells. Biotechnol Adv. 2014;32(6):1133–44.

138. Nioi P, Nguyen T, Sherratt PJ, Pickett CB. The carboxy-terminal Neh3 domain of Nrf2 is required for transcriptional activation. Mol Cell Biol. 2005;25(24):10895–906.

139. Katoh Y, Itoh K, Yoshida E, Miyagishi M, Fukamizu A, Yamamoto M. Two domains of Nrf2 cooperatively bind CBP, a CREB binding protein, and synergistically activate transcription. Genes to Cells. 2001;6(10):857–68.

140. Rada P, Rojo AI, Chowdhry S, McMahon M, Hayes JD, Cuadrado A. SCF/β-TrCP promotes glycogen synthase kinase 3-dependent degradation of the Nrf2 transcription factor in a Keap1-independent manner. Mol Cell Biol. 2011;31(6):1121–33.

141. Wang H, Liu K, Geng M, Gao P, Wu X, Hai Y, et al. RXRα inhibits the NRF2-ARE signaling pathway through a direct interaction with the Neh7 domain of NRF2. Cancer Res. 2013;73(10):3097–108.

142. Kurinna S, Werner S. NRF2 and microRNAs: new but awaited relations. Biochem Soc Trans. 2015;43(4):595–601.

143. Salazar M, Rojo AI, Velasco D, De Sagarra RM, Cuadrado A. Glycogen synthase kinase-3β inhibits the xenobiotic and antioxidant cell response by direct phosphorylation and nuclear exclusion of the transcription factor Nrf2. J Biol Chem. 2006;281:14841–51.

144. Rojo AI, Sagarra MR De, Cuadrado A. GSK-3β down-regulates the transcription factor Nrf2 after oxidant damage: Relevance to exposure of neuronal cells to oxidative stress. J Neurochem. 2008;105:192–202.

145. Wakabayashi N, Skoko JJ, Chartoumpekis D V, Kimura S, Slocum SL, Noda K, et al. Notch-Nrf2 axis: regulation of Nrf2 gene expression and cytoprotection by notch signaling. Mol Cell Biol. 2014;34(4):653–63.

146. Wu T, Zhao F, Gao B, Tan C, Yagishita N, Nakajima T, et al. Hrd1 suppresses Nrf2-mediated cellular protection during liver cirrhosis. Genes Dev. 2014;28(7):708–22.

147. Zimmermann K, Baldinger J, Mayerhofer B, Atanasov AG, Dirsch VM, Heiss EH. Activated AMPK boosts the Nrf2/HO-1 signalling axis - a role for the unfolded protein response. Free Radic Biol Med. 2015;88:1–10.

148. Furukawa M, Xiong Y. BTB protein Keap1 targets antioxidant transcription factor Nrf2 for ubiquitination by the Cullin 3-Roc1 ligase. Mol Cell Biol. 2005;25(1):162–71.

149. Adams J, Kelso R, Cooley L. The kelch repeat superfamily of proteins: propellers of cell function. Trends Cell Biol. 2000;10(1997):17–24.

150. Bryan HK, Olayanju A, Goldring CE, Park BK. The Nrf2 cell defence pathway: Keap1-dependent and -independent mechanisms of regulation. Biochem Pharmacol. 2013;85(6):705–17.

151. Wakabayashi N, Itoh K, Wakabayashi J, Motohashi H, Noda S, Takahashi S, et al. *Keap1*-null mutation leads to postnatal lethality due to constitutive Nrf2 activation. Nat Genet. 2003;35(3):238–45.

152. Devling TWP, Lindsay CD, McLellan LI, McMahon M, Hayes JD. Utility of siRNA against Keap1 as a strategy to stimulate a cancer chemopreventive phenotype. Proc Natl Acad Sci U S A. 2005;102(20):7280–5.

153. Hayes JD, McMahon M. *NRF2* and *KEAP1* mutations: Permanent activation of an adaptive response in cancer. Trends Biochem Sci. 2009;34(4):176–88.

154. Leinonen HM, Kansanen E, Pölönen P, Heinäniemi M, Levonen AL. Role of the Keap1-Nrf2 pathway in cancer. Adv Cancer Res. 2014;122:281–320.

155. Guo Y, Yu S, Zhang C, Tony Kong A-N. Epigenetic regulation of Keap1-Nrf2 signaling. Free Radic Biol Med. 2015;88:337–49.

156. Tebay LE, Robertson H, Durant ST, Vitale SR, Penning TM, Dinkova-Kostova AT, et al. Mechanisms of activation of the transcription factor Nrf2 by redox stressors, nutrient cues, and energy status and the pathways through which it attenuates degenerative disease. Free Radic Biol Med. 2015;88(B):108–46.

157. Suzuki T, Yamamoto M. Molecular basis of the Keap1–Nrf2 system. Free Radic Biol Med. 2015;88:93–100.

158. Hancock R, Schaap M, Pfister H, Wells G. Peptide inhibitors of the Keap1-Nrf2 protein-protein interaction with improved binding and cellular activity. Org Biomol Chem. 2013;11(21):3553–7.

159. Hancock R, Bertrand HC, Tsujita T, Naz S, El-Bakry A, Laoruchupong J, et al. Peptide inhibitors of the Keap1-Nrf2 protein-protein interaction. Free Radic Biol Med. 2012;52(2):444–51.

160. Tong KI, Padmanabhan B, Kobayashi A, Shang C, Hirotsu Y, Yokoyama S, et al. Discovery of a small-molecule inhibitor and cellular probe of Keap1-Nrf2 protein-protein interaction. Bioorganic Med Chem Lett. 2013;23(10):3039–43.

161. Niture SK, Khatri R, Jaiswal AK. Regulation of Nrf2 - An update. Free Radic Biol Med. 2014;66:36–44.

162. Wakabayashi N, Chartoumpekis D V., Kensler TW. Crosstalk between Nrf2 and Notch signaling. Free Radic Biol Med. 2015;88:158–67.

163. Wardyn JD, Ponsford AH, Sanderson CM. Dissecting molecular cross-talk between Nrf2 and

NF-κB response pathways. Biochem Soc Trans. 2015;43(4):621–6.

164. Hayes JD, Chowdhry S, Dinkova-Kostova AT, Sutherland C. Dual regulation of transcription factor Nrf2 by Keap1 and by the combined actions of β-TrCP and GSK-3. Biochem Soc Trans. 2015;43(4):611–20.

165. Namani A, Li Y, Wang XJ, Tang X. Modulation of NRF2 signaling pathway by nuclear receptors: implications for cancer. Biochim Biophys Acta - Mol Cell Res. 2014;1843(9):1875–85.

166. McMahon M, Thomas N, Itoh K, Yamamoto M, Hayes JD. Redox-regulated turnover of Nrf2 is determined by at least two separate protein domains, the redox-sensitive Neh2 degron and the redox-insensitive Neh6 degron. J Biol Chem. 2004;279(30):31556–67.

167. Rada P, Rojo AI, Evrard-Todeschi N, Innamorato NG, Cotte A, Jaworski T, et al. Structural and functional characterization of Nrf2 Degradation by the glycogen synthase kinase 3/β-TrCP axis. Mol Cell Biol. 2012;32(17):3486–99.

168. Chowdhry S, Zhang Y, McMahon M, Sutherland C, Cuadrado A, Hayes JD. Nrf2 is controlled by two distinct β-TrCP recognition motifs in its Neh6 domain, one of which can be modulated by GSK-3 activity. Oncogene. 2013;32(32):3765–81.

169. McMahon M, Itoh K, Yamamoto M, Chanas SA, Henderson CJ, McLellan LI, et al. The cap "n" collar basic leucine zipper transcription factor Nrf2 (NF-E2 p45-related factor 2) controls both constitutive and inducible expression of intestinal detoxification and glutathione biosynthetic enzymes. Cancer Res. 2001;61(8):3299–307.

170. Miao W, Hu L, Scrivens PJ, Batist G. Transcriptional regulation of NF-E2 p45-related factor (NRF2) expression by the aryl hydrocarbon receptor-xenobiotic response element signaling pathway: direct cross-talk between phase I and II drug-metabolizing enzymes. J Biol Chem. 2005;280(21):20340–8.

171. Sanderson LM, Boekschoten M V, Desvergne B, Müller M, Kersten S. Transcriptional profiling reveals divergent roles of PPAR alfa and PPAR beta/delta in regulation of gene expression in mouse liver. Physiol Genomics. 2010;41:42–52.

172. Bell KFS, Al-Mubarak B, Martel M-A, McKay S, Wheelan N, Hasel P, et al. Neuronal development is promoted by weakened intrinsic antioxidant defences due to epigenetic repression of Nrf2. Nat Commun. 2015;6(7066):1–15.

173. Yu S, Khor TO, Cheung K, Li W, Wu T, Huang Y, et al. Nrf2 expression is regulated by epigenetic mechanisms in prostate cancer of TRAMP mice. PLoS One. 2010;5(1):e8579.

174. Sangokoya C, Telen MJ, Chi JT. MicroRNA miR-144 modulates oxidative stress tolerance and associates with anemia severity in sickle cell disease. Blood. 2010;116(20):4338–48.

175. Eades G, Yang M, Yao Y, Zhang Y, Zhou Q. miR-200a regulates Nrf2 activation by targeting Keap1 mRNA in breast cancer cells. J Biol Chem. 2011;286(47):40725–33.

176. Komatsu M, Kurokawa H, Waguri S, Taguchi K, Kobayashi A, Ichimura Y, et al. The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1. Nat Cell Biol. 2010;12(3):213–23.

177. Chen W, Sun Z, Wang XJ, Jiang T, Huang Z, Fang D, et al. Direct interaction between Nrf2 and p21Cip1/WAF1 upregulates the Nrf2-mediated antioxidant response. Mol Cell. 2009;34(6):663–73.

178. Dinkova-Kostova AT, Abramov AY. The emerging role of Nrf2 in mitochondrial function. Free

Radic Biol Med. 2015;88:1–10.

179. Macleod AK, Mcmahon M, Plummer SM, Higgins LG, Penning TM, Igarashi K, et al. Characterization of the cancer chemopreventive NRF2-dependent gene battery in human keratinocytes: Demonstration that the KEAP1-NRF2 pathway, and not the BACH1-NRF2 pathway, controls cytoprotection against electrophiles as well as redox-cycling compounds. Carcinogenesis. 2009;30(9):1571–80.

180. Kitteringham NR, Abdullah A, Walsh J, Randle L, Jenkins RE, Sison R, et al. Proteomic analysis of Nrf2 deficient transgenic mice reveals cellular defence and lipid metabolism as primary Nrf2-dependent pathways in the liver. J Proteomics. 2010;73(8):1612–31.

181. Xu C, Li CY-T, Kong A-NT. Induction of phase I, II and III drug metabolism/transport by xenobiotics. Arch Pharm Res. 2005;28(3):249–68.

182. Nakata K, Tanaka Y, Nakano T, Adachi T, Tanaka H, Kaminuma T, et al. Nuclear receptor-mediated transcriptional regulation in Phase I, II, and III xenobiotic metabolizing systems. Drug Metab Pharmacokinet. 2006;21(6):437–57.

183. Aoki Y, Sato H, Nishimura N, Takahashi S, Itoh K, Yamamoto M. Accelerated DNA adduct formation in the lung of the Nrf2 knockout mouse exposed to diesel exhaust. Toxicol Appl Pharmacol. 2001;173(3):154–60.

184. Ramos-Gomez M, Kwak MK, Dolan PM, Itoh K, Yamamoto M, Talalay P, et al. Sensitivity to carcinogenesis is increased and chemoprotective efficacy of enzyme inducers is lost in Nrf2 transcription factor-deficient mice. Proc Natl Acad Sci U S A. 2001;98(6):3410–5.

185. Chan K, Han XD, Kan YW. An important function of Nrf2 in combating oxidative stress: detoxification of acetaminophen. Proc Natl Acad Sci U S A. 2001;98(8):4611–6.

186. Atilano-Roque A, Wen X, Aleksunes LM, Joy MS. Nrf2 activators as potential modulators of injury in human kidney cells. Toxicol Reports. 2016;3:153–9.

187. Thimmulappa RK, Mai KH, Srisuma S, Kensler TW, Yamamoto M, Biswal S. Identification of Nrf2-regulated genes induced by the chemopreventive agent sulforaphane by oligonucleotide microarray. Cancer Res. 2002;62(18):5196–203.

188. Wu KC, Cui JY, Klaassen CD. Beneficial role of Nrf2 in regulating NADPH generation and consumption. Toxicol Sci. 2011;123(2):590–600.

189. Brachs S, Winkel AF, Polack J, Tang H, Brachs M, Margerie D, et al. Chronic activation of hepatic Nrf2 has no major effect on fatty acid and glucose metabolism in adult mice. PLoS One. 2016;11(11):1–19.

190. Shendure J. The beginning of the end for microarrays? Nat Methods. 2008;5(7):585–7.

191. Yamazaki H, Tanji K, Wakabayashi K, Matsuura S, Itoh K. Role of the Keap1/Nrf2 pathway in neurodegenerative diseases. Pathol Int. 2015;65(5):210–9.

192. Moon EJ, Giaccia A. Dual roles of NRF2 in tumor prevention and progression: Possible implications in cancer treatment. Free Radic Biol Med. 2015;79:292–9.

193. Shin SM, Yang JH, Ki SH. Role of the Nrf2-ARE pathway in liver diseases. Oxid Med Cell Longev. 2013;2013:1–9.

194. Pedruzzi LM, Stockler-Pinto MB, Leite M, Mafra D. Nrf2-Keap1 system versus NF-kB: The good and the evil in chronic kidney disease? Biochimie. 2012;94(12):2461–6.

195. Johnson DA, Johnson JA. Nrf2 - a therapeutic target for the treatment of neurodegenerative diseases. Free Radic Biol Med. 2015;88:253–67.

196. Jakel RJ, Townsend JA, Kraft AD, Johnson JA. Nrf2-mediated protection against 6-hydroxydopamine. Brain Res. 2007;1144(1):192–201.

197. Burton NC, Kensler TW, Guilarte TR. *In vivo* modulation of the Parkinsonian phenotype by Nrf2. Neurotoxicology. 2006;27(6):1094–100.

198. Johnson DA, Amirahmadi S, Ward C, Fabry Z, Johnson JA. The absence of the pro-antioxidant transcription factor Nrf2 exacerbates experimental autoimmune encephalomyelitis. Toxicol Sci. 2009;114(2):237–46.

199. Linker RA, Lee DH, Ryan S, Van Dam AM, Conrad R, Bista P, et al. Fumaric acid esters exert neuroprotective effects in neuroinflammation via activation of the Nrf2 antioxidant pathway. Brain. 2011;134(3):678–92.

200. Pareek TK, Belkadi A, Kesavapany S, Zaremba A, Loh SL, Bai L, et al. Triterpenoid modulation of IL-17 and Nrf-2 expression ameliorates neuroinflammation and promotes remyelination in autoimmune encephalomyelitis. Sci Rep. 2011;1(201):1–11.

201. Ramsey CP, Glass CA, Montgomery MB, Lindl KA, Ritson GP, Chia LA, et al. Expression of Nrf2 in neurodegenerative diseases. Clin Lymphoma. 2007;66(1):75–85.

202. Kanninen K, Malm TM, Jyrkkänen HK, Goldsteins G, Keksa-Goldsteine V, Tanila H, et al. Nuclear factor erythroid 2-related factor 2 protects against beta amyloid. Mol Cell Neurosci. 2008;39(3):302–13.

203. Kubben N, Zhang W, Wang L, Voss TC, Yang J, Qu J, et al. Repression of the antioxidant NRF2 pathway in premature aging. Cell. 2016;165(6):1361–74.

204. Pan H, Guan D, Liu X, Li J, Wang L, Wu J, et al. SIRT6 safeguards human mesenchymal stem cells from oxidative stress by coactivating NRF2. Cell Res. 2016;26(2):190–205.

205. Lau A, Villeneuve NF, Sun Z, Wong PK, Zhang DD. Dual roles of Nrf2 in cancer. Pharmacol Res. 2008;58(5–6):262–70.

206. Taguchi K, Maher JM, Suzuki T, Kawatani Y, Motohashi H, Yamamoto M. Genetic analysis of cytoprotective functions supported by graded expression of Keap1. Mol Cell Biol. 2010;30(12):3016–26.

207. Okawa H, Motohashi H, Kobayashi A, Aburatani H, Kensler TW, Yamamoto M. Hepatocyte-specific deletion of the *keap1* gene activates Nrf2 and confers potent resistance against acute drug toxicity. Biochem Biophys Res Commun. 2006;339(1):79–88.

208. Jaramillo M, Zhang D. The emerging role of the Nrf2–Keap1 signaling pathway in cancer. Genes Dev. 2013;27:2179–91.

209. Ohta T, Iijima K, Miyamoto M, Nakahara I, Tanaka H, Ohtsuji M, et al. Loss of Keap1 function activates Nrf2 and provides advantages for lung cancer cell growth. Cancer Res. 2008;68(5):1303–9.

210. Homma S, Ishii Y, Morishima Y, Yamadori T, Matsuno Y, Haraguchi N, et al. Nrf2 enhances cell proliferation and resistance to anticancer drugs in human lung cancer. Clin Cancer Res. 2009;15(10):3423–32.

211. Gañán-Gómez I, Wei Y, Yang H, Boyano-Adánez MC, García-Manero G. Oncogenic functions of the transcription factor Nrf2. Free Radic Biol Med. 2013;65:750–64.

212. Nioi P, Nguyen T. A mutation of Keap1 found in breast cancer impairs its ability to repress Nrf2 activity. Biochem Biophys Res Commun. 2007;362(4):816–21.

213. Wang XJ, Sun Z, Villeneuve NF, Zhang S, Zhao F, Li Y, et al. Nrf2 enhances resistance of cancer cells to chemotherapeutic drugs, the dark side of Nrf2. Carcinogenesis. 2008;29(6):1235–43.

214. Taguchi K, Motohashi H, Yamamoto M. Molecular mechanisms of the Keap1-Nrf2 pathway in stress response and cancer evolution. Genes to Cells. 2011;16(2):123–40.

215. Hammerman PS, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, et al. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012;489(7417):519–25.

216. Slocum SL, Kensler TW. Nrf2: control of sensitivity to carcinogens. Arch Toxicol. 2011;85(4):273–84.

217. Sporn MB, Liby KT. NRF2 and cancer: the good, the bad and the importance of context. Nat Rev Cancer. 2012;12(8):564–71.

218. Solis LM, Behrens C, Dong W, Suraokar M, Ozburn NC, Moran CA, et al. Nrf2 and Keap1 abnormalities in non-small cell lung carcinoma and association with clinicopathologic features. Clin Cancer Res. 2010;16(14):3743–53.

219. Loignon M, Miao W, Hu L, Bier A, Bismar TA, Scrivens PJ, et al. Cul3 overexpression depletes Nrf2 in breast cancer and is associated with sensitivity to carcinogens, to oxidative stress, and to chemotherapy. Mol Cancer Ther. 2009;8(8):2432–40.

220. Praslicka BJ, Kerins MJ, Ooi A. The complex role of NRF2 in cancer: a genomic view. Curr Opin Toxicol. 2016;1:37–45.

221. Satoh H, Moriguchi T, Takai J, Ebina M, Yamamoto M. Nrf2 prevents initiation but accelerates progression through the KRAS signaling pathway during lung carcinogenesis. Cancer Res. 2013;73(13):4158–68.

222. Mimura J, Itoh K. Role of Nrf2 in the pathogenesis of atherosclerosis. Free Radic Biol Med. 2015;88(Part B):221–32.

223. Barajas B, Che N, Yin F, Rowshanrad A, Orozco LD, Gong KW, et al. NF-E2-related factor 2 promotes atherosclerosis by effects on plasma lipoproteins and cholesterol transport that overshadow antioxidant protection. Arterioscler Thromb Vasc Biol. 2011;31(1):58–66.

224. Freigang S, Ampenberger F, Spohn G, Heer S, Shamshiev AT, Kisielow J, et al. Nrf2 is essential for cholesterol crystal-induced inflammasome activation and exacerbation of atherosclerosis. Eur J Immunol. 2011;41(7):2040–51.

225. Talalay P, Fahey JW, Healy ZR, Wehage SL, Benedict AL, Min C, et al. Sulforaphane mobilizes cellular defenses that protect skin against damage by UV radiation. Proc Natl Acad Sci U S A. 2007;104(44):17500–5.

226. Saw CL, Huang MT, Liu Y, Khor TO, Conney AH, Kong AN. Impact of Nrf2 on UVB-induced skin inflammation/photoprotection and photoprotective effect of sulforaphane. Mol Carcinog. 2011;50(6):479–86.

227. Soeur J, Eilstein J, Lereaux G, Jones C, Marrot L. Skin resistance to oxidative stress induced by resveratrol: from Nrf2 activation to GSH biosynthesis. Free Radic Biol Med. 2015;78:213–23.

228. Schäfer M, Farwanah H, Willrodt AH, Huebner AJ, Sandhoff K, Roop D, et al. Nrf2 links

epidermal barrier function with antioxidant defense. EMBO Mol Med. 2012;4(5):364–79.

229. Schäfer M, Willrodt A, Kurinna S, Link AS, Farwanah H, Geusau A, et al. Activation of Nrf2 in keratinocytes causes chloracne (MADISH)-like skin disease in mice. EMBO Mol Med. 2014;6(4):442–57.

230. Gao Q, Garcia-Pichel F. Microbial ultraviolet sunscreens. Nat Rev Microbiol. 2011;9(11):791–802.

231. Carreto JI, Carignan MO. Mycosporine-like amino acids: relevant secondary metabolites. Chemical and ecological aspects. Mar Drugs. 2011;9(3):387–446.

232. Wada N, Sakamoto T, Matsugo S. Mycosporine-like amino acids and their derivatives as natural antioxidants. Antioxidants. 2015;4:603–46.

233. Oren A, Gunde-Cimerman N. Mycosporines and mycosporine-like amino acids: UV protectants or multipurpose secondary metabolites? FEMS Microbiol Lett. 2007;269:1–10.

234. Sinha RP, Singh SP, Ha D. Database on mycosporines and mycosporine-like amino acids (MAAs) in fungi, cyanobacteria, macroalgae, phytoplankton and animals. J Photochem Photobiol. 2007;89:29–35.

235. Carreto JI, Carignan MO, Montoya NG. Comparative studies on mycosporine-like amino acids, paralytic shellfish toxins and pigment profiles of the toxic dinoflagellates *Alexandrium tamarense*, *A. catenella* and *A. minutum*. Mar Ecol Prog Ser. 2001;223:49–60.

236. Nazifi E, Wada N, Asano T, Nishiuchi T, Iwamuro Y, Chinaka S, et al. Characterization of the chemical diversity of glycosylated mycosporine-like amino acids in the terrestrial cyanobacterium *Nostoc commune*. J Photochem Photobiol B Biol. 2015;142:154–68.

237. D'Agostino PM, Javalkote VS, Mazmouz R, Pickford R, Puranik PR, Neilan BA. Comparative profiling and discovery of novel glycosylated mycosporine-like amino acids in two strains of the cyanobacterium *Scytonema cf. crispum*. Appl Environ Microbiol. 2016;82(19):5951–9.

238. La Barre S, Roullier C, Boustie J. Mycosporine-like amino acids (MAAs) in biological photosystems. In: Stephane La Barre, Kornprobst J-M, editors. Outstanding Marine Molecules: Chemistry, Biology, Analysis. Weinheim, Germany: Wiley-VCH; 2014. p. 333–60.

239. Whitehead K, Hedges JI. Photodegradation and photosensitization of mycosporine-like amino acids. J Photochem Photobiol. 2005;80(2):115–21.

240. Conde FR, Churio MS, Previtali CM. The photoprotector mechanism of mycosporine-like amino acids. Excited-state properties and photostability of porphyra-334 in aqueous solution. J Photochem Photobiol. 2000;56:139–44.

241. Schmid D, Cornelia S, Fred Z. UV-A sunscreen from red algae for protection against premature skin aging. Cosmetics. 2004;139–43.

242. Banaszak AT, Lesser MP, Kuffner IB, Ondrusek M. Relationship between ultraviolet (UV) radiation and mycosporine-like amino-acids (MAA) in marine organisms. Bull Mar Sci. 1998;63(3):617–28.

243. Shick JM, Lesser MP, Dunlap WC, Stochaj WR, Chalker BE, Won JW. Depth-dependent responses to solar ultraviolet radiation and oxidative stress in the zooxanthellate coral *Acropora microphthalma*. Mar Biol. 1995;122(1):41–51.

244. Dunlap WC, Shick JM. Ultraviolet radiation-absorbing mycosporine-like amino acids in coral reef organisms: A biochemical and environmental perspective. J Phycol. 1998;34(3):418–30.

245. Ryu J, Park SJ, Kim IH, Choi YH, Nam TJ. Protective effect of porphyra-334 on UVA-induced photoaging in human skin fibroblasts. Int J Mol Med. 2014;34(3):796–803.

246. Oyamada C, Kaneniwa M, Ebitani K, Murata M, Ishihara K. Mycosporine-like amino acids extracted from scallop (*Patinopecten yessoensis*) ovaries: UV protection and growth stimulation activities on human cells. Mar Biotechnol. 2008;10(2):141–50.

247. Schmid D, Schurch C, Zulli F, Nissen HP, Prieur H. Mycosporine-like amino acids: natural UV-screening compounds from red algae to protect the skin against photoaging. SÖFW-journal. 2003;129(7):38–42.

248. de la Coba F, Aguilera J, de Gálvez M V., Álvarez M, Gallego E, Figueroa FL, et al. Prevention of the ultraviolet effects on clinical and histopathological changes, as well as the heat shock protein-70 expression in mouse skin by topical application of algal UV-absorbing compounds. J Dermatol Sci. 2009;55(3):161–9.

249. Arbeloa EM, Uez MJ, Bertolotti SG, Churio MS. Antioxidant activity of gadusol and occurrence in fish roes from Argentine Sea. Food Chem. 2010;119(2):586–91.

250. Dunlap WC, Yamamoto Y. Small-molecule antioxidants in marine organisms: Antioxidant activity of mycosporine-glycine. Comp Biochem Physiol. 1995;112B(1):105–14.

251. De La Coba F, Aguilera J, Figueroa FL, De Gálvez M V., Herrera E. Antioxidant activity of mycosporine-like amino acids isolated from three red macroalgae and one marine lichen. J Appl Phycol. 2009;21(2):161–9.

252. Suh SS, Hwang J, Park M, Seo HH, Kim HS, Lee JH, et al. Anti-inflammation activities of mycosporine-like amino acids (MAAs) in response to UV radiation suggest potential anti-skin aging activity. Mar Drugs. 2014;12(10):5174–87.

253. Rastogi RP, Sonani RR, Madamwar D, Incharoensakdi A. Characterization and antioxidant functions of mycosporine-like amino acids in the cyanobacterium *Nostoc sp.* R76DM. Algal Res. 2016;16:110–8.

254. Rastogi RP, Incharoensakdi A. UV radiation-induced biosynthesis, stability and antioxidant activity of mycosporine-like amino acids (MAAs) in a unicellular cyanobacterium *Gloeocapsa sp.* CU2556. J Photochem Photobiol B Biol. 2014;130:287–92.

255. Yoshiki M, Tsuge K, Tsuruta Y, Yoshimura T, Koganemaru K, Sumi T, et al. Production of new antioxidant compound from mycosporine-like amino acid, porphyra-334 by heat treatment. Food Chem. 2009;113(4):1127–32.

256. Tao C, Sugawara T, Maeda S, Wang X, Hirata T. Antioxidative activities of a mycosporine-like amino acid, porphyra-334. Fish Sci. 2008;74(5):1166–72.

257. Baldauf SL. Phylogeny for the faint of heart: a tutorial. Trends Genet. 2003;19(6):345–51.

258. Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet. 2005;6(5):361–75.

259. Carrillo H, Lipman D. The multiple sequence alignment problem in biology. SIAM J Appl Math. 1988;48(5):1073–82.

260. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol. 1970;48(3):443–53.

261. Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol. 1981;147(1):195–7.

262. Edgar RC, Batzoglou S. Multiple sequence alignment. Curr Opin Struct Biol. 2006;16(3):368–73.

263. Notredame C. Recent progress in multiple sequence alignment: a survey. Pharmacogenomics. 2002;3(1):131–44.

264. Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S. ProbCons: Probabilistic consistency-based multiple sequence alignment. Genome Res. 2005;15(2):330–40.

265. Wallace IM, O'Sullivan O, Higgins DG, Notredame C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Res. 2006;34(6):1692–9.

266. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91.

267. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.

268. Holder M, Lewis PO. Phylogeny estimation: Traditional and Bayesian approaches. Nat Rev Genet. 2003;4(4):275–84.

269. Sullivan J, Joyce P. Model selection in phylogenetics. Annu Rev Ecol Evol Syst. 2005;36(1):445–66.

270. Jukes TF, Cantor CR. Evolution of protein molecules. In: Munro HN, editor. Mammalian Protein Metabolism. New York: Academic Press, New York; 1969. p. 21–132.

271. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. Vol. 17, American Mathematical Society: Lectures on Mathematics in the Life Sciences. 1986. p. 57–86.

272. Dayhoff MO, Schwartz RM, Orcutt BC. A model of evolutionary change in proteins. Atlas of protein sequence and structure. 1978. p. 345–52.

273. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992;8(3):275–82.

274. Pybus OG. Model selection and the molecular clock. PLoS Biol. 2006;4(5):686–8.

275. E M. Primary Structure and evolution of cytochrome c. Anal Biochem. 1994;217(2):220–30.

276. Bromham L, Penny D. The modern molecular clock. Nat Rev Genet. 2003;4(3):216–24.

277. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56(4):564–77.

278. Chatzou M, Magis C, Chang J-M, Kemena C, Bussotti G, Erb I, et al. Multiple sequence alignment modeling: methods and applications. Brief Bioinform. 2015;17(6):1009–23.

279. Pais FS-M, Ruy P de C, Oliveira G, Coimbra RS. Assessing the efficiency of multiple sequence alignment programs. Algorithms Mol Biol. 2014;9(4):4–8.

280. Martin AP, Palumbi SR. Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci U S A. 1993;90(9):4087–91.

281. Baer CF, Miyamoto MM, Denver DR. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat Rev Genet. 2007;8(8):619–31.

282. Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. Tree of life reveals clock-like speciation and diversification. Mol Biol Evol. 2015;32(4):835–45.

283. Parham JF, Donoghue PCJ, Bell CJ, Calway TD, Head JJ, Holroyd PA, et al. Best practices for justifying fossil calibrations. Syst Biol. 2012;61(2):346–59.

284. Donoghue PCJ, Benton MJ. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. Trends Ecol Evol. 2007;22(8):424–31.

285. McInnes C. Virtual screening strategies in drug discovery. Curr Opin Chem Biol. 2007;11(5):494–502.

286. Shoichet BK. Virtual screening of chemical libraries. Nature. 2004;432(7019):862–5.

287. Broccatelli F, Brown N. Best of both worlds: on the complementarity of ligand-based and structure-based virtual screening. J Chem Inf Model. 2014;54(6):1634–41.

288. Drwal MN, Griffith R. Combination of ligand- and structure-based methods in virtual screening. Drug Discov Today Technol. 2013;10(3):e395–401.

289. A. L, C.D. G. Virtual screening strategies in drug discovery: a critical review. Curr Med Chem. 2013;20(23):2839–60.

290. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2012;64:3–25.

291. Chen B, Harrison RF, Papadatos G, Willett P, Wood DJ, Lewell XQ, et al. Evaluation of machine-learning methods for ligand-based virtual screening. J Comput Aided Mol Des. 2007;21:53–62.

292. Taha M. Mixing pharmacophore modeling and classical QSAR analysis as powerful tool for lead discovery. In: Taha M, editor. Virtual Screening. Shanghai: InTech; 2012. p. 3–15.

293. Vuorinen A, Schuster D. Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. Methods. 2015;71:113–34.

294. Zhuang C, Narayanapillai S, Zhang W, Sham YY, Xing C. Rapid Identification of Keap1−Nrf2 small-molecule inhibitors through structure-based virtual screening and hit-based substructure search. J Med Chem. 2014;57:1121–6.

295. Sun H-P, Jiang Z-Y, Zhang M-Y, Lu M-C, Yang T-T, Pan Y, et al. Novel protein–protein interaction inhibitor of Nrf2–Keap1 discovered by structure-based virtual screening. Medchemcomm. 2014;5(1):93.

296. Lee H, Salzemann J, Jacq N, Chen H, Ho L, Merelli I, et al. Grid-enabled high-throughput *in silico* screening against Influenza A neuraminidase. Comput Syst. 2006;5(4):1–8.

297. Yuriev E, Holien J, Ramsland PA. Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. J Mol Recognit. 2015;28(10):581–604.

298. Yuriev E, Ramsland PA. Latest developments in molecular docking: 2010-2011 in review. J Mol Recognit. 2013;26(5):215–39.

299. Danishuddin M, Khan AU. Structure based virtual screening to discover putative drug candidates: necessary considerations and successful case studies. Methods. 2015;71:135–45.

300. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, et al. The protein data bank: a computer-based archival file for macromolecular structures. Arch Biochem Biophys. 1978;185(2):584–91.

301. Cavasotto CN, Phatak SS. Homology modeling in drug discovery: current trends and applications. Vol. 14, Drug Discovery Today. 2009. p. 676–83.

302. Kaufmann KW, Lemmon GH, Deluca SL, Sheehan JH, Meiler J. Practically useful: what the Rosetta protein modeling suite can do for you. Biochemistry. 2010;49(14):2987–98.

303. Moult J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. Curr Opin Struct Biol. 2005;15(3):285–9.

304. Chen YC. Beware of docking! Trends Pharmacol Sci. 2015;36(2):78–95.

305. Ge H, Wang Y, Li C, Chen N, Xie Y, Xu M, et al. Molecular dynamics-based virtual screening: accelerating the drug discovery process by high-performance computing. J Chem Inf Model. 2013;53(10):2757–64.

306. Good AC, Oprea TI. Optimization of CAMD techniques 3. virtual screening enrichment studies: a help or hindrance in tool selection? J Comput Aided Mol Des. 2008;22(3–4):169–78.

307. Irwin JJ. Community benchmarks for virtual screening. J Comput Aided Mol Des. 2008;22:193–9.

308. Schneider G. Virtual screening: an endless staircase? Nat Rev Drug Discov. 2010;9(4):273–6.

309. Washburn MP. Utilisation of proteomics datasets generated via multidimensional protein identification technology (MudPIT). Brief Funct Genomic Proteomic. 2004;3(3):280–6.

310. Wu CC, MacCoss MJ. Shotgun proteomics: tools for the analysis of complex biological systems. Curr Opin Mol Ther. 2002;4(3):242–50.

311. Florens L, Washburn MP. Proteomic analysis by multidimensional protein identification technology. In: Nedelkov D, Nelson RW, editors. Methods in molecular biology, Vol 328: New and emerging proteomic techniques. Totowa, NJ: Humana Press; 2006. p. 159–75.

312. Philippe Hupé. Mass spectrometry protocol [Internet]. Wikimedia Commons, the free media repository. 2012. Available from: https://commons.wikimedia.org/wiki/File:Mass_spectrometry_protocol.png

313. Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19(3):242–7.

314. Bantscheff M, Lemeer S, Savitski MM, Kuster B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem. 2012;404(4):939–65.

315. Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol. 2010;1–6.

316. Al Shweiki MR, Mönchgesang S, Majovsky P, Thieme D, Trutschel D, Hoehenwarter W. Assessment of label-free quantification in discovery proteomics and impact of technological factors and natural variability of protein abundance. J Proteome Res. 2017;1:1–15.

317. Patel K, Lucas JE, Thompson JW, Dubois LG, Tillmann HL, Thompson AJ, et al. High predictive accuracy of an unbiased proteomic profile for sustained virologic response in chronic hepatitis C patients. Hepatology. 2011;53(6):1809–18.

318. Megger DA, Bracht T, Meyer HE, Sitek B. Label-free quantification in clinical proteomics. Biochim Biophys Acta - Proteins Proteomics. 2013;1834(8):1581–90.

319. Mann M. Functional and quantitative proteomics using SILAC. Nat Rev Mol Cell Biol.

2006;7(12):952–8.

320. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem. 2007;389(4):1017–31.

321. Altelaar AFM, Frese CK, Preisinger C, Hennrich ML, Schram AW, Timmers HTM, et al. Benchmarking stable isotope labeling based quantitative proteomics. J Proteomics. 2013;88:14–26.

322. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem. 2003;75(8):1895–904.

323. Viner R, Bomgarden R, Blank M, Rogers J. Increasing the Multiplexing of Protein Quantitation from 6- to 10-Plex with Reporter Ion Isotopologues. Thermo Scientific; 2013. p. 1–7.

324. Choe KP, Leung CK, Miyamoto MM. Unique structure and regulation of the nematode detoxification gene regulator, SKN-1: implications to understanding and controlling drug resistance. Drug Metab Rev. 2012;44(3):209–23.

325. Sykiotis GP, Bohmann D. Keap1/Nrf2 signaling regulates oxidative stress tolerance and lifespan in *Drosophila*. Dev Cell. 2008;14(1):76–85.

326. Hooper R. Revealing the dawn of photosynthesis. New Sci. 2006;14(2565).

327. Ma Q. Role of Nrf2 in oxidative stress and toxicity. Annu Rev Pharmacol Toxicol. 2013;53(1):401–26.

328. Motohashi H, Yamamoto M. Nrf2-Keap1 defines a physiologically important stress response mechanism. Trends Mol Med. 2004;10(11):549–57.

329. An JH, Vranas K, Lucke M, Inoue H, Hisamoto N, Matsumoto K, et al. Regulation of the *Caenorhabditis elegans* oxidative stress defense protein SKN-1 by glycogen synthase kinase-3. Proc Natl Acad Sci U S A. 2005;102(45):16275–80.

330. Lee J, Godon C, Spector D, Garin J, Toledano MB, Lagniel G, et al. Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast. J Biol Chem. 1999;274(23):16040–6.

331. Zheng M, Storz G. Redox sensing by prokaryotic transcription factors. Biochem Pharmacol. 2000;59(1):1–6.

332. Kim SO, Merchant K, Nudelman R, Beyer WF, Keng T, DeAngelo J, et al. OxyR: a molecular code for redox-related signaling. Cell. 2002;109(3):383–96.

333. Dunlap WC, Fujisawa A, Yamamoto Y. UV radiation increases the reduced coenzyme Q ratio in marine bacteria. Redox Rep. 2002;7(3):3–6.

334. Dunlap WC, Fujisawa A, Yamamoto Y, Inoue M. Tropical UV-tolerant bacteria may provide a pharmacomimetic model for anti-ageing research and cancer prevention. Mar Biotechnol. 2004;6(1):S223–30.

335. Calkins MJ, Johnson DA, Townsend JA, Vargas MR, Dowell JA, Williamson TP, et al. The Nrf2/ARE pathway as a potential therapeutic target in neurodegenerative disease. Antioxid Redox Signal. 2009;11(3):497–508.

336. Chapple SJ, Siow RCM, Mann GE. Crosstalk between Nrf2 and the proteasome: therapeutic potential of Nrf2 inducers in vascular disease and aging. Int J Biochem Cell Biol. 2012;44(8):1315–20.

337. Gao B, Doan A, Hybertson BM. The clinical potential of influencing Nrf2 signaling in degenerative and immunological disorders. Clin Pharmacol Adv Appl. 2014;6(1):19–34.

338. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.

339. Cosconati S, Forli S, Perryman AL, Harris R, Goodsell DS, Olson AJ. Virtual screening with AutoDock: theory and practice. Expert Opin Drug Discov. 2010;5(6):597–607.

340. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, et al. Database resources of the National Center for Biotechnology. Nucleic Acids Res. 2003;31(1):28–33.

341. Federhen S. The NCBI Taxonomy database. Nucleic Acids Res. 2012;40(Database issue):D136–43.

342. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.

343. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–402.

344. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7(10):e1002195.

345. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947–8.

346. Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. Stat Sci. 1992;7(4):457–72.

347. Zmasek C. Archaeopteryx [Internet]. 2014 [cited 2014 Sep 30]. Available from: https://sites.google.com/site/cmzmasek/home/software/archaeopteryx

348. Lo SC, Li X, Henzl MT, Beamer LJ, Hannink M. Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. EMBO J. 2006;25(15):3605–17.

349. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235–42.

350. Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, Brooijmans N, et al. Development and validation of a modular, extensible docking program: DOCK 5. J Comput Aided Mol Des. 2006;20(10–11):601–19.

351. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function , efficient optimization, and multithreading. J Comput Chem. 2009;31(2):455–61.

352. Dunbrack RL. Rotamer libraries in the 21st century. Curr Opin Struct Biol. 2002;12(4):431–40.

353. Huey R, Morris GM, Forli S. Using AutoDock 4 and AutoDock Vina with AutoDockTools: a tutorial [Internet]. 2012. Available from: http://autodock.scripps.edu/faqs-help/tutorial/using-autodock-4-with-autodocktools/2012_ADTtut.pdf

354. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform. 2012;4(17):1–17.

355. Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. J Am Chem Soc. 1992;114(25):10024–35.

356. Adcock SA, McCammon JA. Molecular dynamics: Survey of methods for simulating the activity of proteins. Chem Rev. 2006;106(5):1589–615.

357. ZINC Catalog UEFS Natural Products [Internet]. [cited 2014 May 21]. Available from: http://zinc.docking.org/catalogs/uefsnp

358. ZINC Catalog Nubbe Natural Products [Internet]. [cited 2014 May 21]. Available from: http://zinc.docking.org/catalogs/nubbenp

359. Irwin JJ, Shoichet BK. ZINC - a free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005;45(1):177–82.

360. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605–12.

361. AGN-PC-07CJ71(CID 45378270) - Compound BioActivity Data [Internet]. [cited 2015 Jan 25]. Available from: http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?cid=45378270

362. Xu W, Li X. Antifungal compounds from *Piper* species. Curr Bioact Compd. 2011;7(4):262–7.

363. Gull K, Trinci AP. Griseofulvin inhibits fungal mitosis. Nature. 1973;244(5414):292–4.

364. BouzBouz S, Cossy J. Total synthesis of (+)-strictifolione. Org Lett. 2003;5(11):1995–7.

365. AGN-PC-077JEH | C13H14O4 - PubChem [Internet]. [cited 2015 Jan 23]. Available from: http://pubchem.ncbi.nlm.nih.gov/compound/44575762?from=summary#section=Biological-Test-Results

366. Abdel-Mageed WM, Backheet EY, Khalifa AA, Ibraheim ZZ, Ross SA. Antiparasitic antioxidant phenylpropanoids and iridoid glycosides from *Tecoma mollis*. Fitoterapia. 2012;83(3):500–7.

367. De Marino S, Festa C, Zollo F, Nini A, Antenucci L, Raimo G, et al. Antioxidant activity and chemical components as potential anticancer agents in the olive leaf (*Olea europaea* L. cv Leccino.) decoction. Anticancer Agents Med Chem. 2014;14(10):1376–85.

368. Wang LY, Unehara N, Kitanaka S. Lignans from the roots of *Wikstroemia indica* and their DPPH radical scavenging and nitric oxide inhibitory activities. Chem Pharm Bull (Tokyo). 2005;53(10):1348–51.

369. Da Silva JK, Andrade EHA, Kato MJ, Carreira LM, Guimaraes EF, Maia JGS. Antioxidant capacity and larvicidal and antifungal activities of essential oils and extracts from *Piper krukoffii*. NatProdComm. 2011;6(9):1361–6.

370. Choi JS, Lee HJ, Kang SS. Alaternin, cassiaside and rubrofusarin gentiobioside, radical scavenging principles from the seeds of *Cassia tora* on 1,1-diphenyl-2-picrylhydrazyl (DPPH) radical. Arch Pharm Res. 1994;17(6):462–6.

371. Piao XL, Park IH, Baek SH, Kim HY, Park MK, Park JH. Antioxidative activity of furanocoumarins isolated from *Angelicae dahuricae*. J Ethnopharmacol. 2004;93(2–3):243–6.

372. Jin W, Thuong PT, Su ND, Min BS, Son KH, Chang HW, et al. Antioxidant activity of cleomiscosins A and C isolated from *Acer okamotoanum*. Arch Pharm Res. 2007;30(3):275–81.

373. Li X, Gao Y, Li F, Liang A, Xu Z, Bai Y, et al. Maclurin protects against hydroxyl radical-induced damages to mesenchymal stem cells: antioxidant evaluation and mechanistic insight. Chem Biol Interact. 2014;219:221–8.

374. CHEMBL239211 | C14H10O5 - PubChem [Internet]. [cited 2015 Jan 25]. Available from: http://pubchem.ncbi.nlm.nih.gov/compound/14159818?from=summary#section=BioAssay-Results

375. Epifano F, Genovese S, James Squires E, Gray MA. Nelumal A, the active principle from *Ligularia nelumbifolia*, is a novel farnesoid X receptor agonist. Bioorganic Med Chem Lett. 2012;22(9):3130–5.

376. Rodríguez-Ramiro I, Ramos S, Bravo L, Goya L, Martín MÁ. Procyanidin B2 induces Nrf2 translocation and glutathione S-transferase P1 expression via ERKs and p38-MAPK pathways and protect human colonic cells against oxidative stress. Eur J Nutr. 2012;51(7):881–92.

377. Esatbeyoglu T, Wagner AE, Schini-Kerth VB, Rimbach G. Betanin-A food colorant with biological activity. Mol Nutr Food Res. 2015;59(1):36–47.

378. Li L, Zhang X, Cui L, Wang L, Liu H, Ji H, et al. Ursolic acid promotes the neuroprotection by activating Nrf2 pathway after cerebral ischemia in mice. Brain Res. 2013;1497:32–9.

379. Boettler U, Sommerfeld K, Volz N, Pahlke G, Teller N, Somoza V, et al. Coffee constituents as modulators of Nrf2 nuclear translocation and ARE (EpRE)-dependent gene expression. J Nutr Biochem. 2011;22(5):426–40.

380. Park CH, Noh JS, Kim JH, Tanaka T, Zhao Q, Matsumoto K, et al. Evaluation of morroniside, iridoid glycoside from *Corni fructus*, on diabetes-induced alterations such as oxidative stress, inflammation, and apoptosis in the liver of type 2 diabetic db/db Mice. Biol Pharm Bull. 2011;34(10):1559–65.

381. Kim S, Kim K, Park J, Kwak J, Shik Y, Lee S. Geniposidic acid protects against D-galactosamine and lipopolysaccharide-induced hepatic failure in mice. J Ethnopharmacol. 2013;146(1):271–7.

382. Choi BM, Kim SM, Park TK, Li G, Hong SJ, Park R, et al. Piperine protects cisplatin-induced apoptosis via heme oxygenase-1 induction in auditory cells. J Nutr Biochem. 2007;18(9):615–22.

383. Kong AN, Owuor E, Yu R, Hebbar V, Chen C, Hu R, et al. Induction of xenobiotic enzymes by the MAP kinase pathway and the antioxidant or electrophile response element (ARE/EpRE). Drug Metab Rev. 2001;33(3–4):255–71.

384. Jin X, Liu Q, Jia L, Li M, Wang X. Pinocembrin attenuates 6-OHDA-induced neuronal cell death through Nrf2/ARE pathway in SH-SY5Y cells. Cell Mol Neurobiol. 2015;35(3):323–33.

385. Lee DS, Li B, Kim KS, Jeong GS, Kim EC, Kim YC. Butein protects human dental pulp cells from hydrogen peroxide-induced oxidative toxicity via Nrf2 pathway-dependent heme oxygenase-1 expressions. Toxicol Vitr. 2013;27(2):874–81.

386. Huang CS, Lii CK, Lin AH, Yeh YW, Yao HT, Li CC, et al. Protection by chrysin, apigenin, and luteolin against oxidative stress is mediated by the Nrf2-dependent up-regulation of heme oxygenase 1 and glutamate cysteine ligase in rat primary hepatocytes. Arch Toxicol. 2013;87(1):167–78.

387. Tanigawa S, Fujii M, Hou DX. Action of Nrf2 and Keap1 in ARE-mediated NQO1 expression by quercetin. Free Radic Biol Med. 2007;42(11):1690–703.

388. Podder B, Song H, Kim Y. Naringenin exerts cytoprotective effect against paraquat-induced

toxicity in human bronchial epithelial BEAS-2B cells through NRF2 activation. J Microbiol Biotechnol. 2014;24(5):605–13.

389. Thuong PT, Pokharel YR, Lee MY, Kim SK, Bae K, Su ND, et al. Dual anti-oxidative effects of fraxetin isolated from *Fraxinus rhinchophylla*. Biol Pharm Bull. 2009;32(9):1527–32.

390. Liao W, Ning Z, Chen L, Wei Q, Yuan E, Yang J, et al. Intracellular antioxidant detoxifying effects of diosmetin on 2,2-azobis(2-amidinopropane) dihydrochloride (AAPH)-induced oxidative stress through inhibition of reactive oxygen species generation. J Agric Food Chem. 2014;62(34):8648–54.

391. Bhullar KS, Rupasinghe HPV. Antioxidant and cytoprotective properties of partridgeberry polyphenols. Food Chem. 2015;168:595–605.

392. Barillari J, Canistro D, Paolini M, Ferroni F, Pedulli GF, Iori R, et al. Direct antioxidant activity of purified glucoerucin, the dietary secondary metabolite contained in rocket (*Eruca sativa Mill.*) seeds and sprouts. J Agric Food Chem. 2005;53(7):2475–82.

393. Lee IC, Kim SH, Baek HS, Moon C, Kim SH, Kim YB, et al. Protective effects of diallyl disulfide on carbon tetrachloride-induced hepatotoxicity through activation of Nrf2. Environ Toxicol. 2013;30(5):538–48.

394. Narciso JOA, Soares RODA, Reis Dos Santos Mallet J, Guimarães AE, de Oliveira Chaves MC, Barbosa-Filho JM, et al. Burchellin: study of bioactivity against *Aedes aegypti*. Parasit Vectors. 2014;7(172):1–10.

395. Singh SP, Kumari S, Rastogi RP, Singh KL, Sinha RP. Mycosporine-like amino acids (MAAs): chemical structure, biosynthesis and significance as UV-absorbing/screening compounds. Indian J Exp Biol. 2008;46(1):7–17.

396. Batista JM, Batista ANL, Rinaldo D, Vilegas W, Ambrósio DL, Cicarelli RMB, et al. Absolute configuration and selective trypanocidal activity of gaudichaudianic acid enantiomers. J Nat Prod. 2011;74(5):1154–60.

397. Wang DY, Kumar S, Hedges SB. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc Biol Sci. 1999;266:163–71.

398. Giribet G, Dunn C, Edgecombe G, Rouse G. A modern look at the animal tree of life. Zootaxa. 2007;79:61–79.

399. Oliveira RP, Abate JP, Dilks K, Landis J, Ashraf J, Murphy CT, et al. Condition-adapted stress and longevity gene regulation by *Caenorhabditis elegans* SKN-1/Nrf. Aging Cell. 2009;8(5):524–41.

400. Choe KP, Przybysz AJ, Strange K. The WD40 repeat protein WDR-23 functions with the CUL4/DDB1 ubiquitin ligase to regulate nuclear abundance and activity of SKN-1 in *Caenorhabditis elegans*. Mol Cell Biol. 2009;29(10):2704–15.

401. Wu G, Xu G, Schulman BA, Jeffrey PD, Harper JW, Pavletich NP. Structure of a β-TrCP1-Skp1-β-catenin complex: destruction motif binding and lysine specificity of the SCFβ-TrCP1 ubiquitin ligase. Mol Cell. 2003;11(6):1445–56.

402. Goldstone J V. Environmental sensing and response genes in cnidaria: The chemical defensome in the sea anemone *Nematostella vectensis*. Cell Biol Toxicol. 2008;24(6):483–502.

403. Powell-Coffman JA. Hypoxia signaling and resistance in *C. elegans*. Trends Endocrinol Metab. 2010;21(7):435–40.

404. Sato Y, Itagaki S, Kurokawa T, Ogura J, Kobayashi M, Hirano T, et al. *In vitro* and *in vivo* antioxidant properties of chlorogenic acid and caffeic acid. Int J Pharm. 2011;403(1–2):136–8.

405. Holland HD. The oxygenation of the atmosphere and oceans. Philos Trans R Soc Lond B Biol Sci. 2006;361(1470):903–15.

406. Schirrmeister BE, de Vos JM, Antonelli A, Bagheri HC. Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. Proc Natl Acad Sci U S A. 2013;110(5):1791–6.

407. Lushchak VI. Adaptive response to oxidative stress: bacteria, fungi, plants and animals. Comp Biochem Physiol C Toxicol Pharmacol. 2011;153(2):175–90.

408. Gacesa R, Dunlap WC, Long PF. Bioinformatics analyses provide insight into distant homology of the Keap1-Nrf2 pathway. Free Radic Biol Med. 2015;88(B):373–80.

409. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014;10(4):1–6.

410. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, et al. T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 2011;39(SUPPL. 2):1–5.

411. Chang J-M, Di Tommaso P, Notredame C. TCS: A new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. Mol Biol Evol. 2014;31(6):1625–37.

412. Hedges SB, Marin J, Michael S, Madeline P, Sudhir K. Grand tree of life study shows a clock-like trend in new species emergence and diversity. Mol Biol Evol. 2015;32(4):835–45.

413. Hedges SB, Blair JE, Venturi ML, Shoe JL. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol. 2004;4(2):1–9.

414. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986;3(5):418–26.

415. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000;302(1):205–17.

416. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

417. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.

418. Berner RA, Canfield DE. A new model for atmospheric oxygen over Phanerozoic time. Vol. 289, American Journal of Science. 1989. p. 333–61.

419. Berner RA. GEOCARBSULF : A combined model for Phanerozoic atmospheric O2 and CO2. Geochimica. 2006;70:5653–64.

420. Berner RA. Phanerozoic atmospheric oxygen: new results using the GEOCARBSULF model. Am J Sci. 2009;309(7):603–6.

421. Harada M, Tajika E, Sekine Y. Transition to an oxygen-rich atmosphere with an extensive overshoot triggered by the Paleoproterozoic snowball Earth. Earth Planet Sci Lett. 2015;419:178–86.

422. Lyons TW, Reinhard CT, Planavsky NJ. The rise of oxygen in Earth's early ocean and atmosphere. Nature. 2014;506(7488):307–15.

423. Sperling EA, Halverson GP, Knoll AH, MacDonald FA, Johnston DT. A basin redox transect at the dawn of animal life. Earth Planet Sci Lett. 2013;371–372:143–55.

424. Johnston DT, Poulton SW, Goldberg T, Sergeev VN, Podkovyrov V, Vorobeva NG, et al. Late Ediacaran redox stability and metazoan evolution. Earth Planet Sci Lett. 2012;335–336:25–35.

425. Anbar AD, Duan Y, Lyons TW, Arnold GL, Kendall B, Creaser RA, et al. A whiff of oxygen before the great oxidation event? Science. 2007;317(5846):1903–6.

426. Kopp RE, Kirschvink JL, Hilburn IA, Nash CZ. The Paleoproterozoic snowball Earth: a climate disaster triggered by the evolution of oxygenic photosynthesis. Proc Natl Acad Sci U S A. 2005;102(32):11131–6.

427. Rasmussen B, Fletcher IR, Brocks JJ, Kilburn MR. Reassessing the first appearance of eukaryotes and cyanobacteria. Nature. 2008;455(7216):1101–4.

428. Planavsky NJ, Reinhard CT, Wang X, Thomson D, Mcgoldrick P, Rainbird RH, et al. Low mid-proterozoic atmospheric oxygen levels and the delayed rise of animals. Science. 2014;346(6209):635–8.

429. Mills DB, Canfield DE. Oxygen and animal evolution: did a rise of atmospheric oxygen trigger the origin of animals? BioEssays. 2014;36(12):1145–55.

430. Zhang S, Wang X, Wang H, Bjerrum CJ, Hammarlund EU, Costa MM, et al. Sufficient oxygen for animal respiration 1,400 million years ago. Proc Natl Acad Sci. 2016;113(7):1731–6.

431. Och LM, Shields-Zhou GA. The Neoproterozoic oxygenation event: environmental perturbations and biogeochemical cycling. Earth-Science Rev. 2012;110(1–4):26–57.

432. Van Kranendonk MJ, Altermann W, Beard BL, Hoffman PF, Johnson CM, Kasting JF, et al. A Chronostratigraphic division of the Precambrian: possibilities and challenges. In: Felix MG, G. OggMark James, Schmitz D, Gab MO, editors. The Geologic Time Scale 2012. Boston, MA: Elsevier; 2012. p. 299–392.

433. Shao H, Chu L, Shao M, Jaleel CA, Mi H. Higher plant antioxidants and redox signaling under environmental stresses. C R Biol. 2008;331(6):433–41.

434. Tomitani A, Okada K, Miyashita H, Matthijs HC, Ohno T, Tanaka A. Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts. Nature. 1999;400(6740):159–62.

435. Pitoniak A, Bohmann D. Mechanisms and functions of Nrf2 signaling in *Drosophila*. Free Radic Biol Med. 2015;88:302–13.

436. Blackwell TK, Steinbaugh MJ, Hourihan JM, Ewald CY, Isik M. SKN-1/Nrf, stress responses, and aging in *Caenorhabditis elegans*. Free Radic Biol Med. 2015;88:290–301.

437. Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, et al. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature. 1997;387(6632):489–93.

438. Kuhl M, Cohen Y, Dalsgaard T, Jorgensen BB, Revsbech NP. Microenvironment and photosynthesis of zooxanthellae in scleractinian corals studied with microsensors for O2, pH and light. Mar Ecol Prog Ser. 1995;117(1):159–77.

439. Dunlap WC, Starcevic A, Baranasic D, Diminic J, Zucko J, Gacesa R, et al. KEGG orthology-

based annotation of the predicted proteome of Acropora digitifera: ZoophyteBase - an open access and searchable database of a coral genome. BMC Genomics. 2013;14(1):1–154.

440. Erwin DH, Laflamme M, Tweedt SM, Sperling EA, Pisani D, Peterson KJ. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science. 2011;334(6059):1091–7.

441. Luo ZX, Yuan CX, Meng QJ, Ji Q. A Jurassic eutherian mammal and divergence of marsupials and placentals. Nature. 2011;476(7361):442–5.

442. Luo Z, Martin T. Analysis of molar structure and phylogeny of docodont genera. Bull Carnegie Museum Nat Hist. 2007;39:27–47.

443. Uruno A, Motohashi H. The Keap1-Nrf2 system as an in vivo sensor for electrophiles. Nitric Oxide - Biol Chem. 2011;25(2):153–60.

444. Li X, Zhang D, Hannink M, Beamer LJ. Crystal structure of the Kelch domain of human Keap1. J Biol Chem. 2004;279(52):54750–8.

445. Fukutomi T, Takagi K, Mizushima T, Ohuchi N, Yamamoto M. Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1. Mol Cell Biol. 2014;34(5):832–46.

446. Dinkova-Kostova AT, Holtzclaw WD, Cole RN, Itoh K, Wakabayashi N, Katoh Y, et al. Direct evidence that sulfhydryl groups of Keap1 are the sensors regulating induction of phase 2 enzymes that protect against carcinogens and oxidants. Proc Natl Acad Sci U S A. 2002;99(18):11908–13.

447. Zhang DD, Hannink M. Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative stress. Mol Cell Biol. 2003;23(22):8137–51.

448. Hur W, Gray NS. Small molecule modulators of antioxidant response pathway. Curr Opin Chem Biol. 2011;15(1):162–73.

449. Wells G. Peptide and small molecule inhibitors of the Keap1–Nrf2 protein–protein interaction. Biochem Soc Trans. 2015;43(4):674–9.

450. Jiang ZY, Lu MC, Xu LL, Yang TT, Xi MY, Xu XL, et al. Discovery of potent Keap1-Nrf2 protein-protein interaction inhibitor based on molecular binding determinants analysis. J Med Chem. 2014;57(6):2736–45.

451. Rosic NN. Phylogenetic analysis of genes involved in mycosporine-like amino acid biosynthesis in symbiotic dinoflagellates. Appl Microbiol Biotechnol. 2012;94(1):29–37.

452. Carreto JI, Carignan MO, Montoya NG. A high-resolution reverse-phase liquid chromatography method for the analysis of mycosporine-like amino acids (MAAs) in marine organisms. Mar Biol. 2005;146(2):237–52.

453. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc. 2007;2(9):2212–21.

454. Mishra K, Ojha H, Chaudhury NK. Estimation of antiradical properties of antioxidants using DPPH assay: A critical review and results. Food Chem. 2012;130(4):1036–43.

455. Cardozo KHM, Carvalho VM, Pinto E, Colepicolo P. Fragmentation of mycosporine-like amino acids by hydrogen/deuterium exchange and electrospray ionisation tandem mass spectrometry. Rapid Commun mass Spectrom. 2006;20(2):253–8.

456. Dinkova-Kostova AT, Baird L, Holmström KM, Meyer CJ, Abramov AY. The spatiotemporal regulation of the Keap1–Nrf2 pathway and its importance in cellular bioenergetics. Biochem Soc Trans. 2015;43(4):602–10.

457. Sun Z, Zhang S, Chan JY, Zhang DD. Keap1 controls postinduction repression of the Nrf2-mediated antioxidant response by escorting nuclear export of Nrf2. Mol Cell Biol. 2007;27(18):6334–49.

458. Cheng X, Ku C-H, Siow RCM. Regulation of the Nrf2 antioxidant pathway by microRNAs: New players in micromanaging redox homeostasis. Free Radic Biol Med. 2013;64:4–11.

459. Eggler AL, Liu G, Pezzuto JM, van Breemen RB, Mesecar AD, Breemen RB Van, et al. Modifying specific cysteines of the electrophile-sensing human Keap1 protein is insufficient to disrupt binding to the Nrf2 domain Neh2. Proc Natl Acad Sci U S A. 2005;102(29):10070–5.

460. Abed DA, Goldstein M, Albanyan H, Jin H, Hu L. Discovery of direct inhibitors of Keap1-Nrf2 protein-protein interaction as potential therapeutic and preventive agents. Acta Pharm Sin B. 2015;5(4):285–99.

461. Huang D, Boxin OU, Prior RL. The chemistry behind antioxidant capacity assays. J Agric Food Chem. 2005;53(6):1841–56.

462. Becker K, Hartmann A, Ganzera M, Fuchs D, Gostner JM. Immunomodulatory effects of the mycosporine-like amino acids shinorine and porphyra-334. Mar Drugs. 2016;14(6):1–12.

463. Ryu J, Kwon MJ, Nam TJ. Nrf2 and NF-κB signaling pathways contribute to porphyra-334-mediated inhibition of UVA-induced inflammation in skin fibroblasts. Mar Drugs. 2015;13(8):4721–32.

464. Baird L, Dinkova-Kostova AT. The cytoprotective role of the Keap1-Nrf2 pathway. Arch Toxicol. 2011;85(4):241–72.

465. Mulcahy R, Gipp J. Identification of a putative antioxidant response element in the 5′-flanking region of the human γ-glutamylcysteine synthetase heavy subunit gene. Biochem Biophys Res Commun. 1995;209(1):227–33.

466. Katsuoka F, Motohashi H, Ishii T, Engel JD, Yamamoto M, Aburatani H. Genetic evidence that small Maf proteins are essential for the activation of antioxidant response element-dependent genes. Mol Cell Biol. 2005;25(18):8044–51.

467. Schäfer M, Dütsch S, Auf Dem Keller U, Werner S. Nrf2: a central regulator of UV protection in the epidermis. Cell Cycle. 2010;9(15):2917–8.

468. Kleszczyński K, Ernst IMA, Wagner AE, Kruse N, Zillikens D, Rimbach G, et al. Sulforaphane and phenylethyl isothiocyanate protect human skin against UVR-induced oxidative stress and apoptosis: role of Nrf2-dependent gene expression and antioxidant enzymes. Pharmacol Res. 2013;78:28–40.

469. Dinkova-Kostova AT, Jenkins SN, Fahey JW, Ye L, Wehage SL, Liby KT, et al. Protection against UV-light-induced skin carcinogenesis in SKH-1 high-risk mice by sulforaphane-containing broccoli sprout extracts. Cancer Lett. 2006;240(2):243–52.

470. Hirota A, Kawachi Y, Yamamoto M, Koga T, Hamada K, Otsuka F. Acceleration of UVB-induced photoageing in *nrf2* gene-deficient mice. Exp Dermatol. 2011;20(8):664–8.

471. Engelberg D, Klein C, Martinetto H, Struhl K, Karin M. The UV response involving the Ras signaling pathway and AP-1 transcription factors is conserved between yeast and mammals. Cell. 1994;77(3):381–90.

472. Hinnebusch AG. Translational regulation of *GCN4* and the general amino acid control of yeast. Annu Rev Microbiol. 2005;59(1):407–50.

473. Moye-Rowley WS, Harshman KD, Parker CS. Yeast *YAP1* encodes a novel form of the jun family of transcriptional activator proteins. Genes Dev. 1989;3:283–92.

474. Toledano MB, Delaunay A, Monceau L, Tacnet F. Microbial H2O2 sensors as archetypical redox signaling modules. Trends Biochem Sci. 2004;29(7):351–7.

475. Rodrigues-Pousada C, Menezes RA, Pimentel C. The Yap family and its role in stress response. Yeast. 2010;27:245–58.

476. Zhou BS, Elledge SJ. Checkpoints in perspective. Nature. 2000;408:433–9.

477. Zhou Z, Elledge SJ. *DUN1* encodes a protein kinase that controls the DNA damage response in yeast. Cell. 1993;75(6):1119–27.

478. Pulschen AA, Rodrigues F, Duarte RTD, Araujo GG, Santiago IF, Paulino-Lima IG, et al. UV-resistant yeasts isolated from a high-altitude volcanic area on the Atacama Desert as eukaryotic models for astrobiology. Microbiologyopen. 2015;4(4):574–88.

479. Casteliani AGB, Kavamura VN, Zucchi TD, Sáber ML, Santos R, Frighetto RTS, et al. UV-B resistant yeast inhabit the phyllosphere of strawberry. Br Microbiol Res J. 2014;4(10):1105–17.

480. Corrêa MP, Dubuisson P, Plana-Fattori A. An overview of the ultraviolet index and the skin cancer cases in Brazil. Photochem Photobiol. 2003;78(1):49–54.

481. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25–9.

482. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428(4):726–31.

483. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(421):1–9.

484. Zdobnov EM, Apweiler R. InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17(9):847–8.

485. Punta M, Coggill P, Eberhardt R, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families databases. Nucleic Acids Res. 2012;40:D290–301.

486. Fisher RA. On the interpretation of χ^2 from contingency tables, and the calculation of P. J R Stat Soc. 1922;85(1):87.

487. Griffin TJ. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. Mol Cell Proteomics. 2002;1(4):323–33.

488. MacKay VL, Li X, Flory MR, Turcott E, Law GL, Serikawa KA, et al. Gene expression analyzed by high-resolution state array analysis and quantitative proteomics. Mol Cell Proteomics. 2004;3(5):478–89.

489. Rigbolt KTG, Prokhorova TA, Akimov V, Henningsen J, Johansen PT, Kratchmarova I, et al. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci Signal. 2011;4(164):1–17.

490. Postnikoff SDL, Malo ME, Wong B, Harkness TAA. The yeast forkhead transcription factors

Fkh1 and Fkh2 regulate lifespan and stress response together with the anaphase-promoting complex. PLoS Genet. 2012;8(3):1–13.

491. Martín H, Flández M, Nombela C, Molina M. Protein phosphatases in MAPK signalling: We keep learning from yeast. Mol Microbiol. 2005;58(1):6–16.

492. Choi JH, Lou W, Vancura A. A novel membrane-bound glutathione s-transferase functions in the stationary phase of the yeast *Saccharomyces cerevisiae*. J Biol Chem. 1998;273(45):29915–22.

493. Mooney S, Leuendorf JE, Hendrickson C, Hellmann H. Vitamin B6: A long known compound of surprising complexity. Molecules. 2009;14(1):329–51.

494. Kawamukai M. Biosynthesis and bioproduction of coenzyme Q10 by yeasts and other organisms. Biotechnol Appl Biochem. 2009;53:217–26.

495. Lin MT, Beal MF. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. Nature. 2006;443(7113):787–95.

496. Jamieson DJ. Oxidative stress responses of the yeast *Saccharomyces cerevisiae*. Yeast. 1998;14(16):1511–27.

497. Navarro-Aviño JP, Prasad R, Miralles VJ, Benito RM, Serrano R. A proposal for nomenclature of aldehyde dehydrogenases in Saccharomyces cerevisiae and characterization of the stress-inducible *ALD2* and *ALD3* genes. Yeast. 1999;15(10A):829–42.

498. Altschul AM, Abrams R, Hogness TR. Cytochrome c peroxidase. J Biol Chem. 1940;136(10):777–94.

499. Grant CM. Role of the glutathione/glutaredoxin and thioredoxin systems in yeast growth and response to stress conditions. Mol Microbiol. 2001;39(3):533–41.

500. Lee SM, Koh HJ, Park DC, Song BJ, Huh TL, Park JW. Cytosolic NADP+-dependent isocitrate dehydrogenase status modulates oxidative damage to cells. Free Radic Biol Med. 2002;32(11):1185–96.

501. Beimforde C, Feldberg K, Nylinder S, Rikkinen J, Tuovila H, Dorfelt H, et al. Estimating the Phanerozoic history of the Ascomycota lineages: combining fossil and molecular data. Mol Phylogenet Evol. 2014;77(1):307–19.

502. Parfrey LW, Lahr DJG, Knoll AH, Katz LA. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci U S A. 2011;108(33):13624–9.

503. Rastogi RP, Singh SP, Häder DP, Sinha RP. Detection of reactive oxygen species (ROS) by the oxidant-sensing probe 2',7'-dichlorodihydrofluorescein diacetate in the cyanobacterium *Anabaena variabilis* PCC 7937. Biochem Biophys Res Commun. 2010;397(3):603–7.

504. He Y-Y, Häder D-P. UV-B-induced formation of reactive oxygen species and oxidative damage of the cyanobacterium *Anabaena sp.*: protective effects of ascorbic acid and N-acetyl-l-cysteine. J Photochem Photobiol B Biol. 2002;66(2):115–24.

505. Moline M, Libkind D, del Carmen Dieguez M, van Broock M. Photoprotective role of carotenoids in yeasts: response to UV-B of pigmented and naturally-occurring albino strains. J Photochem Photobiol B Biol. 2009;95(3):156–61.

506. D'Orazio J, Jarrett S, Amaro-Ortiz A, Scott T. UV radiation and the skin. Int J Mol Sci. 2013;14(6):12222–48.

507. Gasch AP, Werner-Washburne M. The genomics of yeast responses to environmental stress and starvation. Funct Integr Genomics. 2002;2:181–92.

508. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, et al. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell. 2000;11(12):4241–57.

509. Villegas LB, Martínez MA, Rodríguez A, Amoroso MJ. Copper resistance and oxidative stress response in *Rhodotorula mucilaginosa* RCL-11. Yeast isolated from contaminated environments in Tucumán, Argentina. In: Bioremediation in Latin America: Current Research and Perspectives. Springer International Publishing Switzerland; 2014. p. 135–48.

510. Gacesa R, Dunlap WC, Barlow DJ, Laskowski RA, Long PF. Rising levels of atmospheric oxygen and evolution of Nrf2. Sci Rep. 2016;6(27740):1–5.

511. Tao S, Justiniano R, Zhang DD, Wondrak GT. The Nrf2-inducers tanshinone I and dihydrotanshinone protect human skin cells and reconstructed human skin against solar simulated UV. Redox Biol. 2013;1(1):532–41.

512. Gems D, Doonan R. Antioxidant defense and aging in *C. elegans*. Cell Cycle. 2009;8(June):1681–7.

513. Schnell N, Krems B, Entian KD. The *PAR1* (*YAP1*/*SNQ3*) gene of Saccharomyces cerevisiae, a c-jun homologue, is involved in oxygen metabolism. Curr Genet. 1992;21(4–5):269–73.

514. Ikner A, Shiozaki K. Yeast signaling pathways in the oxidative stress response. Mutat Res. 2005;569(1–2):13–27.

515. Jindrich K, Degnan BM. The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity. BMC Evol Biol. 2016;16(1):28.

516. Kuge S, Jones N. YAP1 dependent activation of *TRX2* is essential for the response of Saccharomyces cerevisiae to oxidative stress by hydroperoxides. EMBO J. 1994;13(3):655–64.

517. Fernandes L, Rodrigues-Pousada C, Struhl K. Yap, a novel family of eight bZIP proteins in *Saccharomyces cerevisiae* with distinct biological functions. Mol Cell Biol. 1997;17(12):6982–93.

518. Nguyen T, Sherratt PJ, Huang HC, Yang CS, Pickett CB. Increased protein stability as a mechanism that enhances Nrf2-mediated transcriptional activation of the antioxidant response element: degradation of Nrf2 by the 26S proteasome. J Biol Chem. 2003;278(7):4536–41.

519. Alam J, Killeen E, Gong P, Naquin R, Hu B, Stewart D, et al. Heme activates the heme oxygenase-1 gene in renal epithelial cells by stabilizing Nrf2. Am J Physiol Renal Physiol. 2003;284(4):F743–52.

520. Li L, Bagley D, Ward DM, Kaplan J. Yap5 is an iron-responsive transcriptional activator that regulates vacuolar iron storage in yeast. Mol Cell Biol. 2008;28(4):1326–37.

521. Kumar R. An account of fungal 14-3-3 proteins. Eur J Cell Biol. 2017;96(2):206–17.

522. Zhang L, Chen J, Fu H. Suppression of apoptosis signal-regulating kinase 1-induced cell death by 14-3-3 proteins. Proc Natl Acad Sci U S A. 1999;96:8511–5.

523. Yoshimoto H, Saltsman K, Gasch AP, Li HX, Ogawa N, Botstein D, et al. Genome-wide analysis of gene expression regulated by the calcineurin/Crz1p signaling pathway in *Saccharomyces cerevisiae*. J Biol Chem. 2002;277(34):31079–88.

524. Widmann C, Gibson S, Jarpe MB, Johnson GL. Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. Physiol Rev. 1999;79(1):143–80.

525. Inoue H, Hisamoto N, An JH, Oliveira RP, Nishida E, Blackwell TK, et al. The *C. elegans* p38 MAPK pathway regulates nuclear localization of the transcription factor SKN-1 in oxidative stress response. Genes Dev. 2005;19:2278–83.

526. Wang X, Martindale JL, Liu Y, Holbrook NJ. The cellular response to oxidative stress: influences of mitogen-activated protein kinase signalling pathways on cell survival. Biochem J. 1998;333:291–300.

527. Raitt DC, Posas F, Saito H. Yeast Cdc42 GTPase and Ste20 PAK-like kinase regulate Sho1-dependent activation of the Hog1 MAPK pathway. EMBO J. 2000;19(17):4623–31.

528. Tatebayashi K, Yamamoto K, Tanaka K, Tomida T, Maruoka T, Kasukawa E, et al. Adaptor functions of Cdc42, Ste50, and Sho1 in the yeast osmoregulatory HOG MAPK pathway. EMBO J. 2006;25(13):3033–44.

529. Manser E, Leung T, Salihuddin H, Zhao Z, Lim L. A brain serine/threonine protein kinase activated by Cdc42 and Rac1. Nature. 1994;367(6458):40–6.

530. Bagrodia S, Derijard B, Davis RJ, Cerione RA. Cdc42 and PAK-mediated signaling leads to Jun kinase and p38 mitogen-activated protein kinase activation. J Biol Chem. 1995;270(47):27995–8.

531. Bahn Y-S, Kojima K, Cox GM, Heitman J. A unique fungal two-component system regulates stress responses, drug sensitivity, sexual development, and virulence of *Cryptococcus neoformans*. Mol Biol Cell. 2006;17:3122–35.

532. Martins R, Lithgow GJ, Link W. Long live FOXO: Unraveling the role of FOXO proteins in aging and longevity. Aging Cell. 2016;15(2):196–207.

533. Murphy CT, McCarroll S a, Bargmann CI, Fraser A, Kamath RS, Ahringer J, et al. Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. Nature. 2003;424(6946):277–83.

534. Nemoto S, Finkel T. Redox regulation of forkhead proteins through a p66shc-dependent signaling pathway. Science. 2002;295(5564):2450–2.

535. Maoz N, Gabay O, Ben-Asher HW, Cohen HY. The yeast forkhead *HCM1* controls life span independent of calorie restriction. J Gerontol A Biol Sci Med Sci. 2015;70(4):444–53.

536. Rodriguez-Colman MJ, Reverter-Branchat G, Sorolla MA, Tamarit J, Ros J, Cabiscol E. The forkhead transcription factor hcm1 promotes mitochondrial biogenesis and stress resistance in yeast. J Biol Chem. 2010;285(47):37092–101.

537. Devary Y, Gottlieb RA, Smeal T, Karin M. The mammalian ultraviolet response is triggered by activation of Src tyrosine kinases. Cell. 1992;71(7):1081–91.

538. Natarajan K, Meyer MR, Belinda M, Slade D, Roberts C, Alan G, et al. Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. Mol Cell Biol. 2001;21(13):4347–68.

539. Cyert MS. Calcineurin signaling in Saccharomyces cerevisiae: how yeast go crazy in response to stress. Biochem Biophys Res Commun. 2003;311(4):1143–50.

540. Kraus PR, Heitman J. Coping with stress: Calmodulin and calcineurin in model and pathogenic fungi. Biochem Biophys Res Commun. 2003;311(4):1151–7.

541. Ding X, Yu Q, Zhang B, Xu N, Jia C, Dong Y, et al. The type II Ca2+/calmodulin-dependent protein kinases are involved in the regulation of cell wall integrity and oxidative stress response in *Candida albicans*. Biochem Biophys Res Commun. 2014;446(4):1073–8.

542. Ermak G, Davies KJ. Calcium and oxidative stress: from cell signaling to cell death. Mol Immunol. 2002;38(10):713–21.

543. Madeo F, Fröhlich E, Ligr M, Grey M, Sigrist SJ, Wolf DH, et al. Oxygen stress: A regulator of apoptosis in yeast. J Cell Biol. 1999;145(4):757–67.

544. Perrone GG, Tan SX, Dawes IW. Reactive oxygen species and yeast apoptosis. Biochim Biophys Acta - Mol Cell Res. 2008;1783(7):1354–68.

545. Carmona-Gutierrez D, Eisenberg T, Büttner S, Meisinger C, Kroemer G, Madeo F. Apoptosis in yeast: Triggers, pathways, subroutines. Cell Death Differ. 2010;17(5):763–73.

546. Breitenbach M, Laun P, Gimona M. The actin cytoskeleton, RAS–cAMP signaling and mitochondrial ROS in yeast apoptosis. Trends Cell Biol. 2005;15(12):635–7.

547. Braun RJ, Zischka H. Mechanisms of Cdc48/VCP-mediated cell death - from yeast apoptosis to human disease. Biochim Biophys Acta. 2008;1783(7):1418–35.

548. Yamanaka K, Sasagawa Y, Ogura T. Recent advances in p97/VCP/Cdc48 cellular functions. Biochim Biophys Acta. 2012;1823(1):130–7.

549. Ye Y, Meyer HH, Rapoport TA. The AAA ATPase Cdc48/p97 and its partners transport proteins from the ER into the cytosol. Nature. 2001;414(6864):652–6.

550. Wójcik C, Yano M, DeMartino GN, Wojcik C. RNA interference of valosin-containing protein (VCP/p97) reveals multiple cellular roles linked to ubiquitin/proteasome-dependent proteolysis. J Cell Sci. 2004;117(Pt 2):281–92.

551. Tamanoi F. Ras signaling in yeast. Genes Cancer. 2011;2(3):210–5.

552. Heeren G, Jarolim S, Laun P, Rinnerthaler M, Stolze K, Perrone GG, et al. The role of respiration, reactive oxygen species and oxidative stress in mother cell-specific ageing of yeast strains defective in the RAS signalling pathway. FEMS Yeast Res. 2004;5(2):157–67.

553. Gourlay CW, Ayscough KR. Actin-induced hyperactivation of the Ras signaling pathway leads to apoptosis in *Saccharomyces cerevisiae*. Mol Cell Biol. 2006;26(17):6487–501.

554. Phillips AJ, Crowe JD, Ramsdale M. Ras pathway signaling accelerates programmed cell death in the pathogenic fungus Candida albicans. Proc Natl Acad Sci U S A. 2006;103(3):726–31.

555. Downward J. Targeting RAS signalling pathways in cancer therapy. Nat Rev. 2003;3(1):11–22.

556. Vranová E, Inzé D, Van Breusegem F. Signal transduction during oxidative stress. J Exp Bot. 2002;53(372):1227–36.

557. Podda M, Traber MG, Weber C, Yan LJ, Packer L. UV-irradiation depletes antioxidants and causes oxidative damage in a model of human skin. Free Radic Biol Med. 1998;24(1):55–65.

558. He YY, Häder DP. Involvement of reactive oxygen species in the UV-B damage to the cyanobacterium *Anabaena sp.* J Photochem Photobiol B Biol. 2002;66(1):73–80.

559. Shiu CT, Lee TM. Ultraviolet-B-induced oxidative stress and responses of the ascorbate-glutathione cycle in a marine macroalga *Ulva fasciata*. J Exp Bot. 2005;56(421):2851–65.

560. Mackerness SAH, Surplus SL, Jordan BR, Thomas B. Effects of supplementary ultraviolet-B radiation on photosynthetic transcripts at different stages of leaf development and light levels in pea (*Pisum sativum L.*) - role of active oxygen species and antioxidant enzymes. Photochem Photobiol. 1998;68(1):88–96.

561. Davidson JF, Whyte B, Bissinger PH, Schiestl RH. Oxidative stress is involved in heat-induced cell death in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 1996;93(May):5116–21.

562. Shelly C. Lu MD. Glutathione synthesis. Biochim Biophys Acta. 2014;1830(5):3143–53.

563. van Loon AP, Pesold-Hurt B, Schatz G. A yeast mutant lacking mitochondrial manganese-superoxide dismutase is hypersensitive to oxygen. Proc Natl Acad Sci U S A. 1986;83(11):3820–4.

564. Gralla EB. Superoxide dismutase activity is essential for stationary phase survival in *Saccharomyces cerevisiae*. J Biol Chem. 1996;271(21):12275–80.

565. Zhang H, Liu Y, Jiang Q, Li K, Zhao Y, Cao C, et al. Salvianolic acid A protects RPE cells against oxidative stress through activation of Nrf2/HO-1 signaling. Free Radic Biol Med. 2014;69:219–28.

566. Akada R, Shimizu Y, Matsushita Y, Kawahata M, Nishizawa Y, Hoshida H. Use of a *YAP1* overexpression cassette conferring specific resistance to cerulenin and cycloheximide as an efficient selectable marker in the yeast Saccharomyces cerevisiae. Yeast. 2002;19(1):17–28.

567. Yoo HY, Kim SS, Rho HM. Overexpression and simple purification of human superoxide dismutase (SOD1) in yeast and its resistance to oxidative stress. J Biotechnol. 1999;68(1):29–35.

568. Fortuniak A, Zadzinski R, Bilinski T, Bartosz G. Glutathione depletion in the yeast *Saccharomyces cerevisiae*. Biochem Mol Biol Int. 1996;38(5):901–10.

569. Rastogi RP, Richa, Kumar A, Tyagi MB, Sinha RP. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. J Nucleic Acids. 2010;2010:1–32.

570. Ravanat JL, Douki T, Cadet J. Direct and indirect effects of UV radiation on DNA and its components. J Photochem Photobiol B. 2001;63:88–102.

571. Burrows CJ. Surviving an oxygen atmosphere: DNA damage and repair. ACS Symp Ser Am Chem Soc. 2009;2(81):147–56.

572. Sancar GB. Enzymatic photoreactivation: 50 Years and counting. Mutat Res - Fundam Mol Mech Mutagen. 2000;451(1–2):25–37.

573. Stokes MP, Comb MJ. A wide-ranging cellular response to UV damage of DNA. Cell Cycle. 2008;7(14):2097–9.

574. Timson DJ, Singleton MR, Wigley DB. DNA ligases in the repair and replication of DNA. Mutat Res - DNA Repair. 2000;460(3–4):301–18.

575. Boiteux S, Jinks-Robertson S. DNA repair mechanisms and the bypass of DNA damage in *Saccharomyces cerevisiae*. Genetics. 2013;193(4):1025–64.

576. Gadsden MH, McIntosh EM, Game JC, Wilson PJ, Haynes RH. dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae*. EMBO J. 1993;12(11):4425–31.

577. Garbis S, Lubec G, Fountoulakis M. Limitations of current proteomics technologies. J Chromatogr A. 2005;1077(1):1–8.

578. Lindquist S. The heat-shock proteins. Annu Rev Genet. 1988;22:631–78.

579. Papp E, Nardai G, Söti C, Csermely P. Molecular chaperones, stress proteins and redox homeostasis. Biofactors. 2003;17:249–57.

580. Hendrick JP, Hartl F-U. Molecular chaperone functions of heat shock proteins. Ann Rev Biochem. 1993;62:349–84.

581. Frydman J, Nimmesgern E, Erdjument-Bromage H, Wall JS, Tempst P, Hartl F-U. Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. EMBO J. 1992;11(13):4767–78.

582. Kogej T, Gostinčar C, Volkmann M, Gorbushina AA, Gunde-Cimerman N. Mycosporines in extremophilic fungi - novel complementary osmolytes? Environ Chem. 2006;3(2):105–10.

583. Volkmann M, Whitehead K, Rutters H, Rullkotter J, Gorbushina AA. Mycosporine-glutamicol-glucoside: a natural UV-absorbing secondary metabolite of rock-inhabiting microcolonial fungi. Rapid Commun Mass Spectrom. 2003;17(9):897–902.

584. Libkind D, Pérez P, Sommaruga R, Diéguez MDC, Ferraro M, Brizzio S, et al. Constitutive and UV-inducible synthesis of photoprotective compounds (carotenoids and mycosporines) by freshwater yeasts. Photochem Photobiol Sci. 2004;3(3):281–6.

585. Lawrence KP, Long PF, Young AR. Mycosporine-like amino acids for skin photoprotection. Photodermatol Photoimmunol Photomed. (under review).

586. Collinson LP, Dawes IW. Inducibility of the response of yeast cells to peroxide stress. J Gen Microbiol. 1992;138(2):329–35.

587. Wiese AG, Pacifici RE, Davies KJ. Transient adaptation of oxidative stress in mammalian cells. Vol. 318, Archives of biochemistry and biophysics. 1995. p. 231–40.

588. Pickering AM, Vojtovich L, Tower J, Davies KJA. Oxidative stress adaptation with acute, chronic, and repeated stress. Free Radic Biol Med. 2013;55:109–18.

589. Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, et al. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature. 2008;452(7188):745–9.

590. Tong KI, Katoh Y, Kusunoki H, Itoh K, Tanaka T, Yamamoto M. Keap1 Recruits Neh2 through Binding to ETGE and DLG Motifs: Characterization of the Two-Site Molecular Recognition Model. Mol Cell Biol. 2006;26(8):1–14.

591. Hermjakob H, Apweiler R. The Proteomics Identifications Database (PRIDE) and the ProteomExchange Consortium: making proteomics data accessible. Expert Rev Proteomics. 2006;3(1):1–3.

592. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013;41(D1):808–15.

593. Wenger Y, Buzgariu W, Reiter S, Galliot B. Injury-induced immune responses in *Hydra*. Semin Immunol. 2014;26(4):277–94.

594. Stefanik DJ, Friedman LE, Finnerty JR. Collecting, rearing, spawning and inducing regeneration of the starlet sea anemone, *Nematostella vectensis*. Nat Protoc. 2013;8(5):916–23.

595. Quinn B, Gagné F, Blaise C. *Hydra*, a model system for environmental studies. Int J Dev Biol. 2012;52:613–25.

596. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, et al. The dynamic genome of *Hydra*. Nature. 2010;464(7288):592–6.

597. Miller DJ, Ball EE. Cryptic complexity captured: the *Nematostella* genome reveals its secrets. Trends Genet. 2008;24(1):1–4.

598. HGAP Web Portal [Internet]. [cited 2017 Apr 12]. Available from: https://research.nhgri.nih.gov/hydractinia/download/download.cgi?dl=e_assembly

599. Balskus EP, Walsh CT. The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. Science. 2010;329(5999):1653–6.

600. Pope MA, Spence E, Seralvo V, Gacesa R, Heidelberger S, Weston AJ, et al. O-methyltransferase is shared between the pentose phosphate and shikimate pathways and is essential for mycosporine-like amino acid biosynthesis in *Anabaena variabilis* ATCC 29413. ChemBioChem. 2015;16(2):320–7.